DOCUMENT RESUME

ED 274 894 CG 019 398

TITLE	The Evaluator's Reference for Chapter 1 Neglected or Delinquent Youth Programs.
INSTITUTION	RMC Research Corp., Hampton, N.H.
SPONS AGENCY	Office of Vocational and Adult Education (ED), Washington, DC.
PUB DATE	Jan 86
NOTE	230p.; Developed by Region I Technical Assistance Center with other Chapter 1 TACs.
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC10 Plus Postage.
DESCRIPTORS	Adolescents; *Child Neglect; Children; *Delinquency; Elementary Secondary Education; *Evaluation Methods; Program Development; *Program Evaluation; Resource Materials; *Youth Programs

ABSTRACT

This handbook was developed as a resource for individuals involved in the planning and implementation of evaluation activities related to programs for neglected or delinquent youths. The handbook is organized into eight sections. Section 1, Program Evaluation and Overview, discusses a wide range of evaluation purposes, ranging from accountability to determining staff effectiveness. Section 2, Describing the Program, presents 16 possible elements of a description to help the user develop a complete program description, which can be used for fulfilling information needs, for planning activities, and for developing evaluation plans. Section 3, Determining the Focus of Evaluation Activities, describes how the user can identify important program elements and develop specific evaluation questions about the elements. Section 4, Strategies and Techniques for Program Evaluation, presents a variety of techniques for collecting evaluation information. Section 5, Instrumentation, includes an annotated bibliography of pertinent information and characteristics of various norm- and criterion-referenced tests, affective measures, item banks, and other measures which might be appropriate for evaluation. Section 6, Recordkeeping for Program Evaluation and Management, explains different aspects of recordkeeping. Section 7 concerns requirements and recommendations for evaluating Chapter 1 neglected or delinquent programs, and Section 8 presents additional resources for personnel of neglected or delinquent youth programs. Most sections include appendices containing forms, checklists, steps to follow, or detailed descriptions of section topics. (NB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED274894

CORRECTIONS EDUCATION PROGRAM

THE EVALUATOR'S REFERENCE
FOR CHAPTER 1
NEGLECTED OR DELINQUENT

From The

OFFICE OF ELEMENTARY AND SECONDARY EDUCATION

CG 019398

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

January, 1986

CORRECTIONS EDUCATION PROGRAM
Office of Vocational and Adult Education
U.S. Department of Education
Reporters' Building, Room 627
400 Maryland Avenue, S.W.
Washington, D.C. 20202

BEST COPY AVAILABLE

THE EVALUATOR'S REFERENCE FOR CHAPTER 1 NEGLECTED OR DELINQUENT YOUTH PROGRAMS

Developed and Distributed by:
RMC RESEARCH
Region I TAC
400 Lafayette Road
Hampton, NH 03842

April, 1983

INTRODUCTION

This document was originally prepared as a resource for those individuals involved in the planning and implementation of evaluation activities related to N and D programs. Region I Technical Assistance Center, in conjunction with the other Chaper 1 Regional Centers, developed this handbook.

Each office in the Department of Education has a subcommittee on Correctional Education. The purpose of these subcommittees is to work with program staff on concerns and issues of a specific nature to each office and to promote communication, support and delivery of educational services in corrections. One of the activities of the subcommittee is to prepare documents describing the programs and services within each office. It was felt that such a document would be a valuable resource for correctional education programs.

Sincere thanks is extended to Dr. Lawrence Davenport, Assistant Secretary of Elementary and Secondary Education, for his support and promotion of his office staffs' involvement in Correctional Education. Mr. James Evans, Special Assistant, should be recognized for his leadership role as chair of the OESE subcommittee. Specific acknowledgment is also extended to Ms. Delores Hartman who works in this program area and who submitted this document for publication and dissemination.

Additional inquiries may be addressed to:

Ms. Delores Hartman
Chapter I, Western Branch
ROB-3, Room 5114
7th and D Streets, S.W.
Washington, D.C. 20202

(202) 245-2214

We hope that you find this document of value. For information on other available documents or on the Corrections Education Program in the U.S. Department of Education, please contact:

Dr. Dianne Carter
Office of the Assistant Secretary, OVAE
U.S. Department of Education
Reporters Building, Room 627
400 Maryland Avenue, S.W.
Washington, D.C. 20202

(202) 732-2265

TABLE OF CONTENTS

INTRODUCTION: Dr. Dianne Carter
Corrections Education Program
U.S. Department of Education

FORWARD: i-v

CHAPTERS:

1. PROGRAM EVALUATION: AN OVERVIEW

Purposes of Evaluation 1-1

- Accountability and Reporting 1-2
- Determining How Best to Match Services with Individual Student Needs 1-2
- Determining the Degree to Which a Program Has Been Implemented 1-3
- Assessing Short-Term Effects of Programs 1-3
- Assessing Long-Term Effects of Programs 1-3
- Identification and Description of Effective Practices 1-4
- Identification of Relationships Among Services and Program Components 1-4
- Examining Management and Staff Effectiveness 1-4

Situational Constraints 1-5

- Transient Student Populations 1-5
- Institutional Requirements 1-6
- Class Time 1-6
- Achievement Levels 1-6
- Testing Conditions 1-7
- Appropriate Measurement Instruments 1-7
- Procedures for Tracking Students 1-7
- Appropriate Models for Evaluation 1-8

2. DESCRIBING THE PROGRAM

The Relevance of a Complete Program Description 2-1

The Sixteen Program Elements 2-2

- Administration 2-3
- Staffing 2-3
- Budget 2-5
- Student Referral Process 2-6
- Student Selection Criteria 2-6
- Institutional Goals 2-7
- Student and Staff Goals 2-8
- Student and Staff Objectives 2-9
- Developing Individual Student Plans 2-10
- Instructional Activities 2-10
- Coordination Among Courses in the Instructional Program 2-11

Coordination with Follow-up Education and Work Placement 2-12
Student Management Techniques 2-14
Non-Instructional Services 2-14
Recordkeeping 2-15
Program Evaluation 2-16

Appendix 2-A: Program Element Checklist 2-17

3. DETERMINING THE FOCUS OF EVALUATION ACTIVITIES

Introduction 3-1

Selecting Program Elements for Evaluation Purposes 3-1

Determining Evaluation Questions 3-3

Why Evaluation Questions Are Necessary 3-3
Developing the Evaluation Questions 3-4

Appendix 3-A: Ordering Program Elements - Worksheet 3-8
Ordering Program Elements - Example 3-9
Appendix 3-B: Sample Evaluation Questions 3-10

4. STRATEGIES AND TECHNIQUES FOR PROGRAM EVALUATION

Introduction 4-1

Testing Approaches 4-2

Norm-Referenced Testing 4-2
Criterion-Referenced Testing 4-6
Item Banks and Their Uses for Constructing Tests 4-11

Alternative Techniques 4-12

Observations 4-13
Time-on-Task 4-14
Questionnaires 4-16
Interviews 4-18
Existing Records 4-20

Appendix 4-A: Questions to Ask Regarding an Item Bank 4-22
Appendix 4-B: Guidelines for Developing and Conducting Observations 4-25
Appendix 4-C: Guidelines for Developing and Administering Questionnaires 4-27
Appendix 4-D: Guidelines for Developing and Conducting Interviews 4-30
Appendix 4-E: Guidelines for Using Existing Records 4-32

5. INSTRUMENTATION

Introduction 5-1

An Annotated Bibliography 5-1

- List of Annotations 5-2
- Additional Criterion- and Norm-Referenced Tests 5-44
- Additional Inventories - Affective and Other 5-47
- Additional Item Banks 5-47
- Sources for Further Information 5-48

Appendix 5-A: A Criterion-Referenced Test Rating Scale 5-49
Appendix 5-B: A Norm-Referenced Test Rating Scale 5-52
Appendix 5-C: A Suggested Method for Reviewing Tests in Relation to Program Objectives 5-55
Appendix 5-D: Determining When to Test Out-of-Level 5-63
Appendix 5-E: Test Administration Checklist 5-65

6. RECORDKEEPING FOR PROGRAM EVALUATION AND MANAGEMENT

The Relevance of Comprehensive Records 6-1

Types of Records and Their Uses 6-1

- Student Management 6-2
- Short- and Long-Range Planning 6-5
- Evaluation and Administrative Reporting Requirements 6-11

Appendix 6-A: Individual Student File Records 6-14
Appendix 6-B: Individual Student Progress Records 6-20
Appendix 6-C: Group Progress Recording Forms 6-25

7. REQUIREMENTS AND RECOMMENDATIONS FOR EVALUATING CHAPTER 1 N OR D PROGRAMS

Introduction 7-1

Recommended Information to be Collected by Projects 7-1

  Appendix 7-A: Sample Form for Collecting Data for Chapter 1 N or D Projects (Long Version) 7-7
  Appendix 7-B: Sample Form for Collecting Data for Chapter 1 N or D Projects (Short Version) 7-12

8. RESOURCES FOR N OR D PROGRAMS

Introduction 8-1

Technical Assistance Center Services 8-1

  TAC Clearinghouse Resources 8-3
  Available Materials 8-3

The National Diffusion Network 8-4

Appendix 8-A: Technical Assistance Center Regions and Contractors 8-5

FOREWORD

This handbook, The Evaluator's Reference for Chapter 1 Neglected or Delinquent Youth Programs, has been developed as a resource for those individuals involved in the planning and implementation of evaluation activities related to N or D programs. The evaluation process can often be a confusing one, resulting in information that is not of particular use or value to those involved in the day-to-day program activities. This may be even more true for Chapter 1 N or D programs. It is hoped that this handbook will assist the user in planning and conducting evaluations which are feasible and which will provide useful information for program improvement and reporting purposes.

As most people are aware, there has been an intensive movement over the past eight years to improve the quality of evaluation data being reported to Federal and state education agencies, to improve evaluation practices at the state and local level, and to increase the utilization of evaluation information in the improvement of educational programs. ESEA Title I has been the vanguard of this movement through its efforts to establish the Title I Evaluation and Reporting System (TIERS), to encourage Title I programs to follow technically sound guidelines for implementing evaluation models, and to establish regional Technical Assistance Centers to provide free consultative expertise in evaluation to state and local education agencies. While the majority of emphasis was placed on TIERS and evaluation models suitable for Title I programs in mathematics, reading and language arts in grades 2-12, the Department of Education did initiate studies designed to explore whether comparable evaluation models could be developed for Title I Migrant Education, Early Childhood, Non-Instructional and Neglected or Delinquent Programs.

When it became clear that, at least in the case of Title I Neglected or Delinquent Programs, it was not going to be practical or reasonable to develop strict program evaluation models and that in most cases the existing Title I models were not suitable, the Department of Education asked the Region I Title I Technical Assistance Center to lead an effort to develop an evaluation guide or reference for N or D evaluation practices.

The Region I TAC called on assistance from other Title I Technical Assistance Centers across the country who had experience assisting N or D programs, had special expertise in testing or instrumentation or had worked extensively in the area of evaluating program implementation.

This handbook, *The Evaluator's Reference for Chapter 1 Neglected or Delinquent Youth Programs* is a product of the efforts of staff from the:

Region I - TAC	(RMC Research and the University of Rhode Island)
Region II - TAC	(Educational Testing Service in Princeton, NJ)
Region III - TAC	(NTS Raleigh-Durham, NC)
Region IV - TAC	(Educational Testing Service in Atlanta, GA)
Region V - TAC	(Educational Testing Service in Evanston, IL)
Region VIII, IX, X - TAC	(Northwest Regional Educational Lab in Portland, OR)

During a meeting in Washington, D.C. in the late spring of 1982, attended by representatives from the U. S. Department of Education, the Title I TACs and state and local Title I N or D program evaluators and directors, a draft outline for the handbook was developed. From the discussions during this meeting several key points became apparent regarding the handbook and N or D evaluation:

- No model(s) or reporting system would be developed or suggested for N or D program evaluation at this time.
- The unique characteristics of N or D programs and clients made the use of TIERS and existing Title I models inappropriate in most cases.
- Any support document such as this handbook should present sections that address issues related to program implementation and improvement, as well as the reporting of student outcomes and test scores.
- There should be some logical flow to the handbook that would allow a person inexperienced in evaluation an opportunity to address simple, basic issues related to the evaluation of their N or D program and, with assistance from a TAC or other sources, conduct a reasonable program evaluation that would meet their needs and resources.

- Finally, the handbook would be only the first step in helping Chapter 1 N or D programs improve their evaluation practices. In order for this handbook to have maximum effect it would have to be supported by assistance from other resources that would expand on the areas introduced by the handbook.

It is not expected that a user will read through this handbook from cover to cover at one time. Rather, the handbook has been organized so that the user can go directly to any topic of interest (e.g., developing evaluation questions, instrumentation). However, the user can also choose to start at the beginning and systematically develop an N or D program evaluation with little external assistance.

This handbook is organized into eight sections, each of which is briefly described as follows.

1. PROGRAM EVALUATION: AN OVERVIEW

This section discusses the many different purposes for conducting a program evaluation. The intended purpose shapes the entire evaluation, so it must be clarified before undertaking any activities. A wide variety of evaluation purposes is discussed, ranging from accountability to determining staff effectiveness.

2. DESCRIBING THE PROGRAM

This section has been included to help the user develop a complete program description, which can then be used for fulfilling information needs, for planning activities, and for developing evaluation plans. Rather than present a model description, a difficult task because of the variety across N or D programs, 16 possible elements of a description are presented. These program elements, as will be seen, are then used in conjunction with the identified evaluation purpose to plan the actual evaluation.

11

3. DETERMINING THE FOCUS OF EVALUATION ACTIVITIES

This section describes how the user can narrow the evaluation activities first by identifying program elements which are of importance and then by developing specific evaluation questions about the key elements. Because most programs will not have the financial resources and staff time to evaluate everything, it is necessary to determine where to place resources in order to obtain the most useful information for program change and improvement.

4. STRATEGIES AND TECHNIQUES FOR PROGRAM EVALUATION

This section presents a variety of strategies and techniques which can be used to collect evaluation information. Rather than recommend one approach, it is left up to the users to decide which techniques are compatible with their N or D program characteristics, the resources available for the evaluation activities, and the user's preferences toward quantitative or qualitative approaches. Norm- and criterion-referenced testing approaches are discussed, along with alternative data collection techniques, including observations, questionnaires, interviews, and the use of existing records.

5. INSTRUMENTATION

This section provides information that will aid the user in selecting the appropriate instrumentation. An annotated bibliography presents pertinent information and characteristics of various norm- and criterion-referenced tests, affective measures, item banks and other measures which might be appropriate for N or D evaluation activities. Because the bibliography had to be limited in length, a list of additional references to aid in the identification of instrumentation has also been included.

6. RECORDKEEPING FOR PROGRAM EVALUATION AND MANAGEMENT

This section deals with recordkeeping as a critical aspect of the overall evaluation and management of a program. The various types of records and how they can contribute to student management, to short- and long-range planning, and to evaluation and administrative reporting requirements are covered.

12

7. REQUIREMENTS AND RECOMMENDATIONS FOR EVALUATING CHAPTER 1 N OR D PROGRAMS

This section has primarily been included as a place for the user to insert specific regulations and other frequently referenced information. In this way, all of the documentation and information of relevance will be in one easily accessible location. Additionally, an outline of recommended information which might be collected by each program for overall summary purposes is included.

8. RESOURCES FOR N OR D PROGRAMS

This section presents some additional resources which are available to project personnel. The Technical Assistance Center services, including on-site visits, local workshops, telephone consultations, packaged materials, and the Clearinghouse are discussed. The user is also briefly introduced to the National Diffusion Network.

Most of the previously described sections in the handbook also include a variety of appendices. These appendices are typically forms, checklists, steps to be followed or more detailed descriptions of a topic which was introduced in the section. For example, the appendices after Section 5, Instrumentation, include a criterion-referenced test rating scale. Another appendix in this same section summarizes guidelines which help determine when to test out-of-level. These appendices should be duplicated and used as needed.

We hope that this handbook will be a valuable resource for individuals involved in the evaluation of Chapter 1 N or D programs. As new materials become available or existing items change, we will be distributing the upgraded contents. In the meantime, should you have questions or comments about this handbook, please contact your regional Technical Assistance Center.

Everett Barnes, Jr.
Director
Region I, Chapter 1
Technical Assistance Center

1

PROGRAM EVALUATION:
AN OVERVIEW

1. PROGRAM EVALUATION: AN OVERVIEW

PURPOSES OF EVALUATION

Before undertaking the evaluation of a program, one should have a clearly defined purpose in mind. Different purposes suggest different evaluation questions, designs and techniques. The purpose selected really shapes the evaluation and greatly influences the types and uses of evaluation results. In general terms, evaluation can serve to comply with requirements, to find out more about how a program is operating, or to identify effective practices and to improve less effective ones.

Within those broad general categories of intent, there are more specific purposes which might serve as focuses for the evaluation of programs for Neglected or Delinquent youth. Examples of those purposes are:

- Accountability and reporting.
- Determining how best to match services with individual student needs.
- Determining the degree to which a program has been implemented.
- Assessing short-term effects of programs.
- Assessing long-term effects of programs.
- Identification and description of effective practices.
- Identification of relationships among services and program components.
- Examining management and staff effectiveness.

It should be noted that purposes will often overlap and that one evaluation may result in information for more than one purpose.

While not exhaustive, the previous list may help to define the purpose or purposes of an evaluation for a particular program. Each purpose is briefly described as follows, along with some reasons why it might be selected as a focus for evaluation activities.

It is probably not realistic to evaluate all aspects of a program at the same time. By focusing attention on a specific purpose and on one or two aspects or components of a program, an evaluation is more likely to yield useful information. Once an evaluation purpose has been defined and aspects or components of interest have been selected, the next step is the formulation of specific questions which the results of the evaluation should answer. The development of evaluation questions, along with sample questions, is discussed in Section 3.

Accountability and Reporting

It is often necessary to evaluate some aspects of a program to ensure that the program is in compliance with the requirements or expectations of a governing board or funding agency. Usually, the requirements of agencies are based on law and regulations, and frequently the results of such an evaluation must be reported in a specific form on a regular basis.

If accountability and reporting are the primary purposes for evaluating a program, it will probably be necessary to focus the evaluation on determining whether the information needs of the boards or agencies which receive results are being met. Specific reporting requirements will dictate minimum evaluation activities. This type of evaluation usually relies heavily on accurate and complete records. See Section 6 for information on recordkeeping.

Determining How Best to Match Services with Individual Student Needs

Most programs are designed to meet some specific range of student needs. Tailoring program activities to such student needs is a complex, ongoing task. As student populations and needs change, program activities must also change.

One purpose of an evaluation might be to examine or reexamine the range of needs being addressed by the program

16

in light of the needs present in various segments of the student population. Also important for study would be ways in which student needs are identified and how well existing program activities and materials are meeting those needs. This type of evaluation activity might be undertaken if there is concern about instructional effectiveness, the level of student satisfaction, or long-term effectiveness of the program.

Determining the Degree to Which a Program Has Been Implemented

It is not uncommon to find that a program is operating quite differently from the way it was intended to operate. One purpose of an evaluation, therefore, might be to examine how many key elements of the program are actually in place. If an evaluation of the short- or long-term effects of a program is planned, the program's level of implementation should also be evaluated. The results of an evaluation of effects will only be meaningful if there is some assurance that the intended program is in place. An evaluation of program implementation is especially important when new programs are undertaken or new staff are added to the program.

Assessing Short-Term Effects of Programs

Typically an evaluation of short-term effects focuses on changes in students during the course of the program. Such changes may be in cognitive or in other skill areas. Change is generally compared to pre-program behavior or to the behavior of students who are not in the program. This type of evaluation activity can help to point to general program strengths and weaknesses and is often used for accountability purposes.

Assessing Long-Term Effects of Programs

Sometimes the effects of a program are most clearly seen after a student leaves the program. Often much can be learned about a program by examining the experiences of former participants. Programs can also benefit from viewing their effectiveness across several years' operation. When post-program behavior or several years of a program are being studied, the focus of the evaluation is said to be on long-term effects. This type of evaluation activity might be considered when the effects of the program are intended to appear in later job or training performance.

Identification and Description of Effective Practices

Some methods are bound to be more effective than others. An evaluation of short-term or long-term program effects may point to an area of effectiveness which requires further investigation. That investigation could include a review of key program elements and a determination of which elements contribute most to overall program effects.

Evaluating the impact of different techniques and practices on various types of students can produce information which will greatly enhance the effectiveness of a program. It is also important that the dissemination of successful methods be considered, along with use of information from others' experiences.

Identification of Relationships Among Services and Program Components

Often programs are designed and implemented without adequate consideration of how they might be integrated with existing programs. One worthwhile focus for an evaluation might be to examine the interrelationships among programs or program components in order to identify areas of overlap, to redistribute resources, to prevent trapping students in competing or conflicting situations, or to best match programs with students' needs. This type of evaluation activity should be considered if there is confusion about the functions of some components, if key elements are the same across several programs, or if there is difficulty in matching students' needs to programs.

Examining Management and Staff Effectiveness

The effectiveness of a program can sometimes be improved by making better use of staff, enhancing individuals' skills, improving staff communication and morale, or altering management practices. Focusing evaluation activities on this area suggests the assessment of staff strengths and needs and also the assessment of the short-term effects of specific management practices that have been instituted. This type of evaluation activity should be considered when there is friction among staff either in or between programs, when there is difficulty in the implementation of programs, or when personnel changes are frequent.

SITUATIONAL CONSTRAINTS

There are many situations peculiar to an N or D setting which limit the procedures that can be used to evaluate educational programs. It is important to be aware of such situational constraints when planning the evaluation. Although adaptations may be required to help reduce the effect of any constraints, good evaluation procedures can still be practiced in N or D settings. Furthermore, taking existing constraints into consideration before establishing the evaluation procedures will help ensure that the evaluation results are meaningful.

In general, constraints include those situations, regulations or characteristics which nothing can be done about; plans have to be made around constraints. There are, for example, some characteristics of N or D institutions that severely hamper the implementation and evaluation of programs within that setting. There are also problems that can reduce the effectiveness of instruction in the N or D setting. Finally, there is always a variety of miscellaneous constraints which can directly impede any evaluation process. But as long as the existence of certain constraints is known, the evaluation can be planned to reduce their effect. The remainder of this section deals with some constraints which are common to many N or D programs.

Transient Student Populations

Perhaps the most severe limitation to evaluation is the transient nature of the student population. Turnover in N or D settings is often high, resulting in a short duration of instruction. For a variety of reasons, a student's stay in the program may be reduced to a minimum. Before students enter the actual program, they often go through a reception center for observation and testing. Then, once assigned to a program, students may go through an orientation program to familiarize themselves to their new roles and situation. This process further reduces the time a student will spend in the actual program. While the average length of stay varies among settings and from state to state, a national study of N or D programs found that the average student received four months of instruction. Some states, however, have average lengths of stay as short as 2.9 months.

The problem of turnover is compounded by students entering or leaving the program on the basis of institutional needs rather than on educational needs or progress. Often, the instructor has little or no notice that a student is about to enter or leave the program. These types of constraints clearly affect evaluation activities. For example, it becomes more difficult in such situations to implement evaluation activities which require pretest and posttest scores for all students.

Institutional Requirements

Education often has a low priority within N or D programs. Because of this low priority, students often miss class. The students may be needed for work details or may be locked up for disciplinary reasons. In addition, security measures within the institution can interfere with both instruction and evaluation. Evaluation activities must be planned to take into account any specific institutional restrictions.

Class Time

Compounding the problem of a short time spent in the program, two other problems further reduce instructional time: absenteeism and less than full use of class time. Absenteeism may occur for many reasons, such as student illness, disciplinary measures, institutional requirements, or rehabilitation (such as visits to the rehabilitation worker). Less than full use of class time may also occur if the instructor is absent and class is cancelled because a substitute is not available -- or a substitute may be present, but without adequate lesson plans, resulting in misused time.

Even with the students and instructors present, class time is often consumed by non-instructional activities. For example, one national study found that 37 percent of class time in N or D institutions was spent on non-instructional activities. The evaluation must be planned to take these constraints into consideration. The use of class time may even become the focus of the evaluation (see the discussion of time-on-task in Section 4).

Achievement Levels

Another set of problems can result when the students are functioning far below their age expectancy. N or D students often have a history of failure, thus student motivation tends to be low. Since there is a lack of high interest-low ability materials to teach basic education skills, the students often use materials developed for younger students. This lack of appropriate materials can, in turn, have a negative effect on student progress.

Testing Conditions

Testing conditions may be less than ideal, especially if tests are given during the first few weeks the student is in the institution. Testing may take place in a diagnostic center where the student has already been given a series of other tests. Also, the student is probably not in the best frame of mind at this time -- a factor which may further contribute to an inaccurate measure of ability. In addition, it has been noted that some students, familiar with the system, will suppress their test scores so that they will be placed in special programs or so that they can more easily show educational progress.

Appropriate Measurement Instruments

Most of the tests being used for evaluation of programs in N or D settings were designed for use with average students in non-institutional settings. Test norms often do not extend to adolescents who are functioning at lower educational levels. There are, however, alternatives to using commonly available standardized tests. There are also some instruments available which measure progress in adult basic education curricula. (For more information on assessment instruments, see Section 5.)

Another constraint in regard to testing is that standardized tests are frequently developed to measure progress from fall to spring. Many tests sample broad bands of achievement and are not sensitive enough to measure progress in the brief periods of time N or D students participate in some programs. Furthermore, the test norms have been developed for specific times during the year; for test scores to be meaningful, the N or D student must usually take the test during the same period. Any variations can affect the evaluation results.

Procedures for Tracking Students

Students' previous school records can be difficult to obtain, if they are accessible at all. Often the students have been out of school for quite a while and so have no records. The time it takes to track down any records is also a problem for the many students who stay in the program for a short time. Post-release information necessary to evaluate the long-term effectiveness of programs is equally hard to obtain. Parole officers are often the best equipped to gather follow-up information about students who have left the program, but large caseloads may prevent them from doing so.

Appropriate Models for Evaluation

Clearly an overriding problem is the lack of evaluation models that take into account the various constraints discussed. For that reason, the evaluation of N or D programs requires more flexibility in selecting and implementing evaluation techniques. The remaining sections of this handbook are designed to suggest some possible evaluation alternatives. Since there are no easily applicable evaluation models available, procedures which can be used for the evaluation of programs in spite of the possible constraints in N or D settings will be discussed.

2

DESCRIBING THE PROGRAM

2. DESCRIBING THE PROGRAM

THE RELEVANCE OF A COMPLETE PROGRAM DESCRIPTION

Developing a complete and accurate description of an operating program is not a simple task. It requires an unbiased view, careful attention to detail and an understanding of the role each part of the program plays within the whole system. It is not surprising then that few educational programs have routinely developed program descriptions which extend beyond an abstract or a list of program elements.

A complete program description is useful in several ways. First, it is a vehicle for clear communication about the program both externally and internally. A description serves the information needs of funding and governing agencies or boards; accountability begins with a common understanding of the intentions of program designers. In a similar way, a complete program description can serve internal staff communication needs as well. A written guide to the program answers questions of new and old staff in a definite manner that can be supplemented with, but not replaced by, the collective oral history of staff members. Potential problems are prevented by the existence of clear procedures. Decision-making about new issues is facilitated when all parties can make reference to common information. The potential for developing creative solutions to problems is increased when staff members can spend less time and energy reinventing policies and procedures as they are implementing the program.

Second, complete program descriptions are invaluable in the planning process. Certainly, if a new program is being undertaken, clear descriptions of all major elements will facilitate implementation. Intelligent planning for resource allocation with expanding or contracting budgets demands accurate descriptions of program intentions and an understanding of how all elements function. Coordination across different programs serving Neglected or Delinquent students is one of the more difficult tasks faced by program managers. Locating and eliminating areas in which programs overlap and identifying gaps in services requires detailed descriptions of program services and functions.

Finally, a complete program description is the basis for the development of evaluation plans which yield useful results. The development of evaluation questions (see Section 3) is grounded in a knowledge of the intentions and operation of program elements. Further, a thorough description of program elements is necessary to frame realistic recommendations for program modification based on the results of evaluations.

THE SIXTEEN PROGRAM ELEMENTS

It is clear then that a good program description will serve a variety of important functions. Because N or D programs vary greatly, it is not possible to develop a model description that would be applicable to most programs. Instead, this section includes guidelines for developing descriptions of sixteen major program elements which would be common to many projects. Each element is discussed and the components of a description are listed.

It is unlikely that any one program description would include a lengthy narrative about each element. The previous discussion of purposes suggests that one might develop a description of an element if any of these conditions were true:

- An evaluation focusing on certain program areas is being planned (e.g., a study of time spent on instructional areas).
- There is concern or confusion about an element (e.g., no one is sure about student selection procedures).
- Changes in an element have been suggested or are planned (e.g., a new testing procedure is to be inaugurated).
- Major staff or administration changes are about to occur.
- Outside support is being sought for a program.

Another option is to build a complete program description over time, selecting a few elements to describe thoroughly during each program year. This can be done by first using the Program Element Checklist (Appendix 2-A) to review current program descriptions for completeness. On the basis of the results of the checklist, additional program element descriptions can be developed where needed.

1. Administration

Administrative activities include those policies, procedures and routines that are required to operate programs and services effectively. It is important to specify administrative activities so that expectations are clearly known to all staff and coordination among program elements is possible. Clarity of expectations and coordination ensure smooth program operations. A description of administration includes:

- procedures for supervision and evaluation of staff
- procedures for ensuring effective communication of information to staff and to others (funding agency, other services, etc.)
- hierarchy of reporting relationships
- budget authorization policies
- staff recruitment and hiring practices
- policies for grievance and conflict resolution related to students and staff
- long-range and short-range planning procedures
- establishing a documentation system that supports compliance with regulations
- procedures which ensure the health and safety of staff and students
- procedures for obtaining and upgrading staff benefits
- administrative roles and responsibilities

2. Staffing

The element of staffing encompasses four areas: (1) staffing patterns, which includes job categories and student-staff ratios; (2) responsibilities for each job category; (3) staff background and qualifications; and (4) staff development. The delivery of any program depends in large part on the quality of staff; quality can be enhanced by providing sufficient numbers of staff, selecting those with appropriate experience and backgrounds, providing adequate support services and carefully matching backgrounds with tasks required. A description of staffing patterns includes:

- types of job categories (e.g., instructor, aides, counselor, coordinator)
- numbers of full and part-time staff employed in each category
- overall student-staff ratios
- specific program student-staff ratios (e.g., vocational training, basic skills)

A description of responsibilities for job categories includes minimum expectations for performance related to:

- instructional, administrative, and non-instructional duties
- recordkeeping, evaluation and reporting
- maintenance and upgrading of professional skills
- student management tasks
- special committee assignments (e.g., providing input for hiring and review of other staff)

A description of staff background and qualifications includes:

- a summary of experiential and academic preparation of staff members by job category
- unique requirements of job responsibilities and how staff characteristics match those requirements
- a summary of special interests and abilities

A description of staff development includes:

- the process by which staff training needs are identified
- in-service activities planned to meet those needs
- evaluation procedures for in-service activities
- examples of training needs which have been identified
- examples of the types of in-service activities which are conducted

3. Budget

A budget description includes the dollar amounts to be expended for specific program purposes. Detailed budget descriptions provide both the proof that resources have been allocated to match major goals of the program and the guidance required to make choices among competing expenditure alternatives. Include in a budget description:

- the costs projected by line-item category (e.g., equipment, salaries for instruction, instructional supplies)
- the costs projected by general purpose (e.g., academic activities, vocational training, counseling program)
- a projection of costs per student
- a description of costs which are one-time expenditures (e.g., program start-up costs, facilities)
- authorization to spend policies
- a description of budget planning procedures
- internal fiscal accounting procedures
- fiscal reporting schedule

4. Student Referral Process

Student referral is the process by which students who are potentially eligible for a program are identified. A well-defined referral process ensures that those students most in need and whose needs are most in accord with the program become the eligible candidates for the next steps in the process. A well-defined process produces better referrals than a casual system. It also ensures that students who may not seem appropriate for the program under consideration are referred to another service or program. A description of the student referral process includes:

- designation of who refers students (staff members or outside agencies such as other institutions, the courts)
- the means by which referring individuals receive sufficient information about the program upon which to base a referral (e.g., written or oral program descriptions, memos, individuals who are responsible for generating referrals from others)
- approximate time schedule for receipt of referrals
- content of the referral, including whatever judgments or information must be provided by referring individuals
- any forms used in the process

5. Student Selection Criteria

Student selection criteria are the formal standards by which students are accepted into a program or service. It is important to remember that criteria do change as programs and services develop; criteria should reflect needs assessment findings and the programs designed to meet those needs. A clear description of student selection criteria ensures compliance with regulations and allows a determination of whether standards are fair and appropriate. Well-defined criteria ensure that students in the program are those whose needs are most in accord with the program. A complete description of the student selection criteria includes:

- a description of the relevant characteristics of the target audience as determined by needs assessments (e.g., achievement and demographic information)
- the process by which a pool of eligible students is identified
- the procedure by which the most educationally needy within that pool are rank-ordered
- the selection indicator(s) used in the above procedure (e.g., tests, instructor referral, self-referral)
- means of obtaining information for indicator(s)
- the way in which indicators are combined to identify each student's degree of need (composite, multiple cut-offs, single criteria)
- the actual cut-off scores used (e.g., 25th percentile, 65 out of 100 points on a composite score, or three of five indicators of need show eligibility)
- waiting list policy for filling open slots
- exit criteria which specify expectations of success or mastery

6. Institutional Goals

Institutional goals are statements which describe the desired outcomes obtainable by the program as a whole. These goals may encompass a wide variety of areas, such as: upgrading or adding new facilities; restructuring the program to meet projected changes in student needs; establishing new funding sources; or strengthening positive community attitudes toward the program. Institutional goals are typically more long-term in nature and may require the involvement of a variety of program personnel. A description of institutional goals should include statements specifying:

- the goal and how reaching it will enhance the program
- a timeline for attaining each goal, including the various tasks to be performed

- who is responsible for coordinating the activities required to reach each goal
- an identification of other personnel who will be involved in the activities
- a way to evaluate whether the goal has actually been reached

7. Student and Staff Goals

Student and staff goals are broad statements of the outcomes obtainable by students and staff through planned activities. Each goal may include a variety of cognitive or skill outcomes and may be either short-term or long-term in nature. Typically, one goal will encompass a group of measurable objectives. All or only some of the goals may be attained depending upon interests, abilities and time available. A description of goals should include statements specifying:

- goals categorized by target audience
- who is responsible for developing new goals and revising existing ones
- where written copies of all goals are filed
- how goals are used to plan instructional objectives, activities, and purchase of commercial materials
- how goals are used to develop individual student plans
- who receives copies of goals and how this information is used by recipients (i.e., potential employers, counselors)
- how goals are used in initial test selection activities
- how goals are used to monitor overall student progress
- how goals are used to monitor staff development
- how goals are used in evaluation activities

8. Student and Staff Objectives

Student and staff objectives are measurable statements of the outcomes obtainable through planned activities. As with goals, objectives may be cognitive or in other skill areas. The mastery of one or more objectives should lead to attainment of specific goals. Objectives may cover a fairly broad range of outcomes or may focus on a single, rather narrow outcome.

Objectives should always be stated in measurable terms and include the following three parts: (1) the conditions -- a statement which describes the circumstances under which the outcome will be demonstrated; (2) the performance -- a statement which includes one or more measurable verbs which describe the outcome; and, (3) the criterion -- a statement which specifies the minimum acceptable standard which must be reached in order to demonstrate mastery of the objective. A description of objectives should include statements specifying:

- objectives categorized by target audience and goals
- who is responsible for developing new objectives and revising existing ones
- where written copies of all objectives are filed
- how objectives are used to develop, revise, or adapt activities
- how objectives are used to organize the content and sequence of a program or course
- how objectives are used to develop individual student plans
- who receives copies of objectives and how this information is used by recipients (i.e., potential employers, counselors, follow-up educational program personnel)
- how objectives are used to develop or select tests
- how objectives are used to monitor student progress
- how objectives are used to coordinate activities among courses within the institution

- how objectives are used to coordinate activities with follow-up and work placement outside of the institution
- how objectives are used in evaluation activities

9. Developing Individual Student Plans

Developing individual student plans involves the identification of goals and objectives most suitable to meet each student's needs and then selecting or developing activities which will ultimately help the student reach the desired outcomes. A variety of information can contribute to the development of a plan, including diagnostic tests, achievement tests, interest surveys, affective surveys, counseling sessions, background information, learning style preference surveys, and discussions with the student. Descriptions of individual student plans should specify:

- standard format for each plan
- where plans are stored and how confidentiality is ensured
- who is responsible for obtaining information to be used in the plan
- how and where various types of information are obtained
- when plans will be developed and updated
- how plans are revised in response to unscheduled occurrences (such as a student not progressing or the availability of new information)
- how plans are revised in response to planned growth and development (such as goals being met)
- who is involved in the development of a plan (e.g., past instructors, present instructors, counselors, parents, guardians, the student)

10. Instructional Activities

Instructional activities include all of the planned learning experiences, media, and hardware used to teach objectives.

Within any setting, there should be a variety of deliberately planned activities -- group discussion, simulations, print materials, filmstrips and more -- each of which has as its purpose to teach the student a specific topic, skill, or attitude. Instructional activities may include those developed by an instructor, commercial materials or a combination of items.

Ideally, instructional activities are organized around specific objectives. Furthermore, in order to better meet individual needs, there should be more than one approach available to teach any one objective. For example, students who have difficulty reading would benefit from instruction that is not dependent upon reading skills; audiotapes or high interest-low readability texts would be more effective than standard textbooks. Or, since some students may require more repetition and practice than others, instructional activities with additional practice exercises would be of benefit.

Descriptions of instructional activities should be developed for all aspects of a program, including academic, vocational, and real world survival skills. These descriptions should specify:

- activities cross-referenced to goals and objectives
- commercial and staff-developed programs in use
- how staff and students access materials
- preferred instructional approaches
- examples of activities
- how staff work together to develop, adapt, and revise activities
- orientation of new staff to instructional approaches
- predominant types of equipment available
- specific facilities employed

11. Coordination Among Courses in the Instructional Program

Coordination among courses in the instructional program is necessary in order to provide a more cohesive and integrated sequence of instruction. Just as it is necessary to carefully and logically order goals and objectives within a course, the same care must be taken across courses offered within a program. Ensuring that this coordination is done allows instructors to eliminate unnecessary overlap across courses and at the same time to develop course sequences by which students can move successfully from an entry level to a more advanced course. While a brief review may be appropriate, a complete instructional sequence over previously learned topics would be a non-effective use of both instructor and student time. On the other hand, if an advanced course assumes that certain entry level skills have already been taught, failure to teach those skills would mean that the instructor of the more advanced course would have to spend time providing instruction that should have been taught elsewhere.

Coordination across courses can also be used to increase student motivation and progress. For example, students who are very interested in their auto mechanics course may perform better in a basic math course if the math skills are taught in the context of mechanics. Similarly, reading skills might develop faster and with less resistance if taught through a driver's license manual or a repair manual. However, this coordination will not always occur unless it is deliberately planned; thus the plans for coordination encourage communication and ultimately result in a more successful experience for the students involved.

A description of the process for coordination among courses should specify:

- the plan for ensuring this coordination
- who is responsible for organizing, monitoring, and conducting coordination activities
- how individuals are identified for participation in coordination activities
- when coordination activities occur
- the process for updating and revising linkages

12. Coordination with Follow-up Education and Work Placement

Coordination with follow-up education and work placement is just as critical as coordination within the program. Although it is realized that coordination with follow-up placements is often not possible, it is desirable and so is discussed in this section. Because the ultimate purpose of any program is student success at the next step, each instructional plan should be developed with the next placement in mind. The student will be more capable of success at this next step if the instructor is aware of the necessary entry-level capabilities. For example, if the student will be placed in a public high school work-study program, mastery of certain cognitive and affective skills will be critical for success. If a student is entering a vocational training program, a somewhat different set of skills may be necessary. A student who will be going directly into a job placement situation will have very different needs.

The individuals responsible for administering the follow-up education or work program will be able to identify specific cognitive or other skill areas which are entry-level capabilities for their particular program. For example, a high school mathematics instructor will be best able to identify the entry-level mathematics skills. The job placement counselor will be more sensitive to behavior such as correct dress and appropriate interview skills. The parole officer will be able to identify specific skills which will help the student adjust to new situations. Coordinating and communicating with these types of people will not only help ensure student success but will also make them more willing to accept the students into their particular program; they will have a better understanding of the students' capabilities.

A description of the process for coordination with follow-up education and work placement should specify:

- the plan for ensuring this coordination
- who is responsible within the institution for organizing, monitoring, and conducting coordination activities
- how individuals within the institution are identified for participation in coordination activities
- the process for identifying the follow-up education and work placement programs that should be linked to internal programs
- the process for identifying and contacting other outside support systems
- when coordination activities occur
- the process for updating and revising linkages

13. Student Management Techniques

Student management techniques encompass the wide range of methods used to encourage growth on the part of the student in cognitive and other skill areas. In addition to providing the student with activities which teach specific objectives, instructors typically employ management techniques which motivate the student to learn. These techniques range from non-verbal reinforcements such as a smile to very formalized token rewards for prespecified behaviors. Formal contracts, grades, and modeling are other examples of management techniques.

Ideally, the management techniques used should be matched to the student; different individuals are motivated in different ways. For example, a younger student may be encouraged by a positive non-verbal gesture or positive verbal approval from the instructor. In contrast, an older student with a history of academic failure and poor interpersonal skills may do better if a more concrete system of rewards, such as accumulating points to buy free time, is implemented. A variety of management techniques may need to be tried before the most effective ones are identified. If possible, the management techniques used for each student should be consistent from instructor to instructor. Descriptions of management techniques used should specify:

- how management techniques are designed for students who are involved
- examples of techniques presently being used
- institution-wide standard practices
- schedule of periodic review for effectiveness of techniques

14. Non-Instructional Services

Non-instructional services include those program components primarily designed to promote health, safety, and other non-cognitive goals for students. Activities may or may not be directly related to instruction. Types of services which might be included are individual and group counseling, follow-up work and education placements, preventive medical and dental care, recreational and leisure time activities, career placement, and family service activities.

It is especially important to develop complete descriptions of non-instructional services because their functions are not likely to be well understood by students, institutional staff not directly involved in the service, and those external to the institution, including funding agencies. Descriptions will help to ensure that duplication of services does not occur, that student referrals are appropriate, and that all students can take advantage of opportunities available through institutions. A description of each non-instructional service includes:

- goals and objectives or statements of purpose
- examples of activities and materials
- numbers and types of staff
- numbers and characteristics of students involved
- recordkeeping procedures
- evaluation activities

15. Recordkeeping

Recordkeeping procedures are a critical aspect of the overall management of a program, especially in situations in which students may enter or leave a program at variable times during the year and in which students may come into the program with a wide range of backgrounds, capabilities and interests. Comprehensive records can actually be the key to a variety of activities, including: the coordination of individual and group progress through both the instructional and non-instructional activities in the program; the conducting of short- and long-range planning activities to develop a program most responsive to student needs; the conducting of evaluation activities in order to determine program effectiveness; and the meeting of administrative reporting requirements.

Comprehensive records should include the following types of information: individual student files; individual student progress records; group progress records; a list of goals, objectives and activities cross-referenced by target audience; an inventory of consumable and non-consumable resources; a list of outside resources and types of services offered; staff files; follow-up academic, vocational, and job placement opportunity; student attendance records; student selection procedures and documentation; results of past surveys administered to staff or students; follow-up data collected on students in academic, vocational and job placements; interview documentation with potential employers; results of past evaluation activities; state reports; and all other information which may form a basis for future planning or evaluation activities.

A description of recordkeeping procedures should include:

- an identification of the types of records maintained and the individual(s) who are responsible for collecting and updating the necessary information
- an explanation of how the various records are used by students, instructors, administrators and other program or outside personnel
- samples of recordkeeping forms, where appropriate

16. Program Evaluation

Program evaluation is the process of systematically gathering information to determine the value or effectiveness of program elements and services. Evaluation can serve a variety of purposes. Section 1 of this document provides examples of evaluation purposes (e.g., determining short-term and long-term effects of projects, establishing the degree to which projects have been implemented). Further, evaluation activities can be of many different types (see Section 4 for descriptions of evaluation techniques especially for N or D projects). Evaluations are typically designed to answer a set of questions which are considered important to staff -- questions which affect long- and short-term program development and operations. The evaluation results should form the basis for improving programs and services. All evaluation activities should always be described before any evaluation activity actually begins. The worth of evaluation is in its use. Maximum use of information requires the systematic and careful collection and examination of information as well as the involvement of those who will be affected by the results. A good description of evaluation plans is essential so that the results will be comprehensive, comprehensible, credible and useful. A complete description of the evaluation of each program includes:

- a statement of the purpose of evaluation
- key evaluation questions
- management of the evaluation
- evaluation design

Appendix 2-A

PROGRAM ELEMENT CHECKLIST

The first step in developing a complete program description is to review, for completeness, the information currently available in written form about each major program element. The Program Element Checklist was designed to facilitate this review process. The checklist is organized by the sixteen major program elements which were discussed in Section 2. Each major element is further subdivided into specific topics.

- To use the checklist, rate the current status of documentation about the topic under consideration. If a description exists and is both adequate and accurate, place a check mark in the column labeled YES.

- If the topic is either not described in written form, or if existing descriptions are inadequate or inaccurate, place a check mark in the column labeled NO.

- For any topics where a NO has been checked, use the column labeled NOTES to indicate what needs to be done in the way of additional documentation or modification in order for the description of the topic to be complete. Use the NOTES column to also indicate what partial information exists, if any, where further information might be located, or who might be involved in developing the description.

- If the topic is not a part of the program or is not applicable in some way, place a check mark in the column labeled NA.

Once the checklist has been completed, decisions can be made about which elements should be further described by reviewing the items checked NO.

Person Completing

Checklist ________________

Date ________________

Appendix 2-A

PROGRAM ELEMENT CHECKLIST

YES - indicates that the written description is adequate

NO - indicates inadequacy, lack of clarity, or lack of documentation

NA - indicates that this aspect of the element is not relevant

ADMINISTRATION	YES	NO	NA	NOTES
1. Procedures for staff supervision				
2. Policies for staff evaluation				
3. Procedures for ensuring effective communication of information to staff				
4. Procedures for ensuring effective communication of information to others				
5. Hierarchy of reporting relationships				
6. Policies for budget authorization				
7. Procedures for staff recruitment and hiring				
8. Policies for grievance and conflict resolution for students				
9. Policies for grievance and conflict resolution for staff				
10. Procedures for long-range planning				

ADMINISTRATION	YES	NO	NA	NOTES
11. Procedures for short-range planning				
12. Documentation system that supports compliance with regulations				
13. Procedures to ensure health and safety of staff				
14. Procedures to ensure health and safety of students				
15. Policies for obtaining and upgrading staff benefits				
16. Administrative roles and responsibilities				

STAFFING	YES	NO	NA	NOTES
1. Job categories				
2. Numbers of full and part-time staff employed in each category				
3. Overall student-staff ratios				
4. Staff-student ratios by specific programs				
5. Instructional responsibilities for each job category				
6. Administrative responsibilities for each job category				
7. Non-instructional responsibilities for each job category				

STAFFING	YES	NO	NA	NOTES
8. Recordkeeping responsibilities for each job category				
9. Evaluation and reporting responsibilities for each job category				
10. Procedures for maintaining and upgrading of professional skills for each job category				
11. Student management tasks for each job category				
12. Special committee assignments for each job category				
13. Experiential and academic preparation of staff by each job category				
14. Unique requirements of job responsibilities and how staff characteristics match those requirements				
15. Special interests and abilities of staff by each job category				
16. Procedures for identifying staff training needs				
17. In-service activities to meet staff training needs				
18. Procedures for evaluating in-service training activities				
19. Examples of training needs which have been identified				
20. Examples of types of in-service activities which are conducted				

BUDGET	YES	NO	NA	NOTES
1. Costs projected by line-item category				
2. Costs projected by general purpose				
3. Projection of costs per student				
4. Description of costs which are one-time expenditures				
5. Policies for authorizing budget allocations				
6. Procedures for planning budgets				
7. Procedures for internal fiscal accounting				
8. Schedule for fiscal reporting				

STUDENT REFERRAL PROCESS	YES	NO	NA	NOTES
1. Individuals who refer students				
2. Procedures by which referring individuals receive information about program upon which to base a referral				
3. Approximate time schedule for receipt of referrals				
4. Content of the referral				
5. Forms which are used in the referral process				

STUDENT SELECTION CRITERIA	YES	NO	NA	NOTES
1. Relevant characteristics of the target audience as determined by needs assessments				

STUDENT SELECTION CRITERIA	YES	NO	NA	NOTES
2. Procedures by which a pool of eligible students is identified				
3. Procedures by which the most educational needy within that pool are rank-ordered				
4. Selection indicators used in the procedures				
5. Means of obtaining information for the indicators				
6. Way in which indicators are combined to identify each student's degree of need				
7. Cut-off scores used				
8. Policies for filling open slots in the program from the waiting list				
9. Exit criteria which specify expectations of success or mastery				

INSTITUTIONAL GOALS	YES	NO	NA	NOTES
1. Goal and how reaching it will enhance the program				
2. Timeline for attaining each goal, including various tasks to be performed				
3. Individual responsible for coordinating activities required to reach each goal				
4. Other personnel who will be involved in the activities				

INSTITUTIONAL GOALS	YES	NO	NA	NOTES
5. Procedures for evaluating whether the goal has been reached				

STUDENT AND STAFF GOALS	YES	NO	NA	NOTES
1. Goals categorized by target audience				
2. Individual(s) responsible for developing new goals and revising existing ones				
3. Where written copies of all goals are filed				
4. Procedures for using goals to plan instructional objectives, activities, and purchase of commercial materials				
5. Procedures for using goals to develop individual student plans				
6. Individuals who receive copies of goals and how this information is used by them				
7. Procedures for using goals in initial test selection activities				
8. Procedures for using goals to monitor overall student progress				
9. Procedures for using goals to monitor staff development				
10. Procedures for using goals in evaluation activities				

STUDENT AND STAFF OBJECTIVES	YES	NO	NA	NOTES
1. Objectives categorized by target audience and goals				
2. Individual(s) responsible for developing new objectives and revising existing ones				
3. Where written copies of all objectives are filed				
4. Procedures for using objectives to develop, revise, or adapt activities				
5. Procedures for using objectives to organize content and sequence of a program or course				
6. Procedures for using objectives to develop individual student plans				
7. Individuals who receive copies of objectives and how this information is used by them				
8. Procedures for using objectives to develop or select tests				
9. Procedures for using objectives to monitor student progress				
10. Procedures for using objectives to coordinate activities among courses within the institution				
11. Procedures for using objectives to coordinate activities with follow-up and work placement outside of the institution				
12. Procedures for using objectives in evaluation activities				

DEVELOPING INDIVIDUAL STUDENT PLANS	YES	NO	NA	NOTES
1. Standard format for each plan				
2. Where plans are stored and how confidentiality is ensured				
3. Individual responsible for obtaining information to be used in the plan				
4. Procedures for obtaining various types of information and timeline for doing so				
5. Policies for when plans will be developed and updated				
6. Procedures for revising plans in response to unscheduled occurrences				
7. Procedures for revising plans in response to planned growth and development				
8. Individual(s) involved in the development of a plan				

INSTRUCTIONAL ACTIVITIES	YES	NO	NA	NOTES
1. Activities cross-referenced to goals and objectives				
2. Commercial and staff-developed programs in use				
3. Procedures by which staff and students obtain materials				
4. Preferred instructional approaches				

INSTRUCTIONAL ACTIVITIES	YES	NO	NA	NOTES
5. Examples of activities				
6. Procedures for how staff work together to develop, adapt, and revise activities				
7. Procedures for orientation of new staff to instructional approaches				
8. Predominant types of equipment available				
9. Specific facilities employed				

COORDINATION AMONG COURSES IN THE INSTRUCTIONAL PROGRAM	YES	NO	NA	NOTES
1. Procedures for ensuring coordination				
2. Individual responsible for organizing, monitoring, and conducting coordination activities				
3. Process for identifying individuals for participation in coordination activities				
4. Schedule of coordination activities				
5. Process for updating and revising linkages				

COORDINATION WITH FOLLOW-UP EDUCATION AND WORK PLACEMENT	YES	NO	NA	NOTES
1. Procedures for ensuring coordination				

COORDINATION WITH FOLLOW-UP EDUCATION AND WORK PLACEMENT	YES	NO	NA	NOTES
2. Individual responsible within the institution for organizing, monitoring, and conducting coordination activities				
3. Process for identifying follow-up education and work placement programs that should be linked to internal programs				
4. Process for identifying and contacting other outside support systems				
5. Schedule of coordination activities				
6. Process for updating and revising linkages				

STUDENT MANAGEMENT TECHNIQUES	YES	NO	NA	NOTES
1. Process by which management techniques are designed for students				
2. Examples of techniques presently being used				
3. Institution-wide standard practices				
4. Schedule of periodic review for effectiveness of techniques				

NON-INSTRUCTIONAL SERVICES	YES	NO	NA	NOTES
1. Goals and objectives or statements of purpose				
2. Examples of activities and materials				

NON-INSTRUCTIONAL SERVICES	YES	NO	NA	NOTES
3. Numbers and types of staff involved				
4. Numbers and characteristics of students involved				
5. Procedures for maintaining records on non-instructional services				
6. Procedures for evaluating non-instructional services				

RECORDKEEPING	YES	NO	NA	NOTES
1. Types of records maintained and individual(s) responsible for collecting and updating the necessary information				
2. How the various records are used by students, instructors, administrators and other program or outside personnel				
3. Samples of recordkeeping				

PROGRAM EVALUATION	YES	NO	NA	NOTES
1. Purpose of the evaluation activities				
2. Key evaluation questions				
3. Process for management of the evaluation				
4. Evaluation design				

3. DETERMINING THE FOCUS OF EVALUATION ACTIVITIES

INTRODUCTION

Evaluation activities should be focused first by clarifying the purposes of the evaluation and then by identifying the specific program elements to be addressed. Most programs will have neither the financial resources nor the staff time to conduct yearly evaluations of all of their elements. The problem, then, is to determine where to apply the resources available for evaluation purposes. Obviously, the evaluation activities required by Federal, state and local agencies must be conducted; but what other aspects of a program should be evaluated? How is a determination made as to the placement of resources necessary to obtain the most useful information for program change and improvement?

There are two steps which, if followed, will help to determine where to best use resources and to guide in the development of an evaluation plan. First, program elements of importance should be identified. Then, specific evaluation questions should be developed for each program element of interest. This section presents a process for accomplishing these two steps. The results of this process will be one or more well-focused evaluation questions -- questions which, when answered, will provide the staff with information for program improvement.

SELECTING PROGRAM ELEMENTS FOR EVALUATION PURPOSES

In order to select elements to be evaluated, the key program elements must first be identified. Complete descriptions of elements, such as those discussed in Section 2, are a reference point for this identification. The choice of which elements to evaluate can then be made on an informal basis, by identifying those of greatest concern or those about which little is known. Or, if it is unclear as to which program elements should be evaluated, a more systematic process can be used to help guide the determination. One such brief process which will result in an ordered list of elements is as follows.

With the existing program description as a guide, a list of program elements can be developed. Each element should be paraphrased clearly and succinctly and then rated by asking the four questions provided. Each response should then be scored as suggested below.

Question 1: Has this element been evaluated before?

NO - score 2 (two) points
SOMEWHAT - score 1 (one) point
YES - score 0 (zero) points

Question 2: Will evaluating this element provide information that will help to make decisions or policies?

YES - score 2 (two) points
MAYBE - score 1 (one) point
PROBABLY NOT - score 0 (zero) points

Question 3: Has this element been of concern or problematic in some way?

VERY MUCH - score 2 (two) points
SOMEWHAT - score 1 (one) point
NOT MUCH - score 0 (zero) points

Question 4: Have there been external requests for information about this element?

YES, DIRECT REQUESTS - score 2 (two) points
YES, BUT INDIRECTLY OR IMPLIED - score 1 (one) point
NONE - score 0 (zero) points

Now, for each element, the score for each of the four questions is added up to arrive at a total score for that element. This number is then used to rank order the elements. Those with the highest score are probably most relevant for the present evaluation purposes.

Following this process will result in an ordered list of program elements -- a list which indicates where best to use the resources available for evaluation activities. Rather than using limited resources to attempt to evaluate everything, which often results in superficial information, it is better to conduct a thorough evaluation of one or two program elements during the course of the year and then move on to other program elements in following years.

Appendix 3-A includes a worksheet which can be used to order program elements, following the steps and process just described. Also included is a brief example showing how this procedure worked for one program.

DETERMINING EVALUATION QUESTIONS

After the program elements have been selected for evaluation purposes, it is necessary to determine the questions which will be answered. Developing specific questions will help focus the evaluation activities on producing useful information which directly addresses the actual topics of concern.

Why Evaluation Questions Are Necessary

Consider briefly what might happen if specific questions are not developed before conducting the evaluation activities. Take a situation, for example, where the decision has been made to evaluate students' cognitive achievement through the use of a standardized achievement test. After pre- and posttests are administered, the students' average gains are calculated. The results are then presented to the members of an advisory committee and the program instructors. Unfortunately, however, during the meeting the committee members say, "But what we really want to know is: What specific skills have been mastered? . . . How many students actually mastered each skill?" The instructors, on the other hand, ask: "What specific skills are students not mastering? In what areas does instruction need to be improved?"

Some of these unanswered questions may be answered by rescoring tests or reanalyzing data, but this would require additional staff time. Or, the answers to these questions might not be available at all because the test used did not collect the necessary information. In any case, by not first determining a specific evaluation question or set of questions, the results might be unsuitable for meeting the real needs of those using the information.

This need for determining the questions becomes even more critical when evaluation activities move further away from the use of the conventional achievement test and more toward the collection of information for program implementation or process evaluation purposes. For example, in a situation where one of the major program elements to be evaluated is that of instructor-student interactions, what types of interactions should be measured: verbal, non-verbal, or both? Should evaluation activities focus on interactions oriented to group management, interpersonal relationships, direct instruction, or all three? These types of decisions depend on what the staff and policymakers want to know.

The possibilities for uncertainty regarding what to focus on multiply rapidly in less conventional evaluation areas. These uncertainties, such as choosing the type of instrumentation, data collection techniques, analysis procedures and so forth, need to be resolved as much as possible before beginning the evaluation activities. Having well specified evaluation questions before beginning any activities helps ensure that the evaluation will proceed without wasting time, that information will not have to be collected again, and that the results will be useful for program evaluation purposes.

Developing the Evaluation Questions

After one or more program elements have been selected for the evaluation activities, it is time to develop the actual questions. Each element to be evaluated should be considered one at a time in order to develop a list of questions. These questions should be ones whose answers will provide information on the extent to which this element is effective or on how successfully it has been implemented.

For example, if the evaluation is focusing on the remedial mathematics element of a program, some questions whose answers may help evaluate that element might include:

1. At what level are the students' math skills on entering the program, in comparison with their peers not in the program?
2. How does student participation in math activities compare with participation in reading activities?
3. To what extent do the instructors believe that the program is effective for students' math learning?
4. To what extent do the instructors think that students are appropriately selected for the program?

5. Do instructors diagnose student needs accurately?

6. Do instructors prescribe instructional objectives and learning activities which directly address student needs?

7. Does the program contribute to students' doing well in later programs?

The answers to each of these questions could provide information useful for determining the success of the program or the level of program implementation.

There is no formula for ensuring that the evaluation questions developed are the best ones. However, there are three guidelines which, if followed, will make it more likely that the questions asked will provide useful evaluation information in an efficient manner. These guidelines are clarity, relevance and utility.

For a question to follow the guideline of clarity, the meaning of that question must be clear to those who read the question; they should be able to agree on what the question means. There are two areas which are often unclear and thus special attention should be paid to them when writing the questions: (1) the object to be measured, and (2) the standard or comparison for that measure. With most evaluation questions there will be a standard or comparison, but in some situations the question may only describe something; thus there will not be a standard or comparison.

Consider the issue of clarity in regard to the evaluation question "How effective is the math component?" What is the object that will be measured? What standard will be used to measure this effectiveness? This evaluation question is too vague and thus could mean many things to different readers. In comparison, consider the second question in the list previously given: "How does student participation in math activities compare with participation in reading activities?" It is fairly clear that the objects to be measured are those of student participation in math and reading activities. Further, in this case a comparison of the level of participation in math activities to the level of participation in reading activities will be used to determine effectiveness.

Consider briefly the fifth question in the list: "Do instructors diagnose student needs accurately?" The object to be measured here is how well the instructors do at diagnosing student needs. To answer this question it would be necessary to devise some measure of instructors' abilities to diagnose those needs, perhaps by having each instructor prepare a written documentation of a diagnosis for a particular student. A model diagnosis could be developed to act as the standard, with another person making a judgment as to how well each instructor had diagnosed the needs. By comparing instructors' diagnoses with the model it would be possible to determine how well the teachers actually do at diagnosing student needs.

For an evaluation question to follow the guideline of .relevance, each question should (1) contribute information useful in fulfilling the intended purpose of the evaluation activities, and (2) provide useful information to those who determine program policy and to those who implement the program. It is important that the purpose of the evaluation activities, as discussed in Section 1, should be already identified before this point. Questions which will not provide information toward the intended purpose, although they may be very interesting, should probably not be pursued unless resources are unlimited. Appendix 3-B includes a list of evaluation purposes, as discussed in Section 1, along with some sample evaluation questions. These questions are provided only as examples and certainly do not cover the full range of evaluation questions that may be asked in relation to a specific project.

For the guideline of relevance to be fully met, the questions should also provide useful information to those who actually implement the program. The questions asked should provide information to the project director, the members of the advisory board, the program staff, and others involved with the program. The best way to ensure that the questions are relevant is to involve these people in formulating the questions during the planning activities for the evaluation.

For an evaluation question to follow the guideline of utility, there must be a projection about whether those with appropriate authority will really use, or be open to using, the results. For example, consider the evaluation question: "To what extent are the goals of the program sufficiently focused to guide instructional, supportive, and student assessment activities?" If it is already known that a Board member, such as a State Education Department staffer, and the Program Director already have concerns about this issue, then they are likely to be especially attuned to answers to the question. They will be more likely to consider the information seriously in changing policies. Therefore, if the question is also clear and relevant, it should get high priority for evaluation activities.

In comparison, consider the evaluation question: "In what ways do the interactions of the Project Director and the staff affect the instructional activities?" If those planning the evaluation realize that the Project Director may have problems with the staff, is very resistant to criticism or change, and is secure in his or her position, then answers to these types of questions will probably not be used constructively. Therefore, the question is probably not a very practical one and might be eliminated; more constructive or useful questions should probably be asked.

Once potential evaluation questions have been reviewed for clarity, relevance and utility, a final concept should be addressed: the questions should cover a broad range of aspects for the element being evaluated. In other words, the questions should not focus on a very narrow aspect of the program element, such as only attendance data or only cognitive outcomes. There is a broad range of possible questions for each element and this range should be fully covered.

Consider the question discussed previously: "How does student participation in math activities compare with participation in reading activities?" One might simply collect basic factual information on who did or did not participate in math and reading activities. Or participation could be dealt with on another level where some interpretation by the instructor or another observer must be made about the extent of a student's involvement in an activity beyond mere presence. This would be at a different level than the attendance, in the sense that it addresses more complex patterns of behavior. The resulting information would probably provide more useful information for program improvement than would a simple record of attendance.

Question 5, "Do instructors diagnose needs accurately?" is another example of the type of question that requires more than the basic recording of factual information. A judgment is required about a fairly complex pattern of behavior on the part of the instructor. Finally, consider the question: "Does the program contribute to students' doing well in follow-up programs?" This question moves the evaluation activities to another level: determining whether the program helps students in follow-up placements.

It should not be implied that some types of questions are better than others for evaluation purposes. In fact, sometimes the answers to basic factual questions must be collected in order to interpret answers to other questions. The point here is simply that the evaluation questions should focus on a variety of types of questions and not just on one type.

Appendix 3-A

ORDERING PROGRAM ELEMENTS - WORKSHEET

ELEMENT	Evaluated Before No - 2 Somewhat - 1 Yes - 0	Information Will Help Decisions/ Policies Yes - 2 Maybe - 1 Probably Not - 0	Problematic/High Concern Area: Very Much - 2 Somewhat - 1 Not Much - 0	External Request For Information: Yes, Direct - 2 Yes, Indirect - 1 None - 0	TOTAL

ORDERING PROGRAM ELEMENTS - EXAMPLE

ELEMENT	Evaluated Before No - 2 Somewhat - 1 Yes - 0	Information Will Help Decisions/ Policies Yes - 2 Maybe - 1 Probably Not - 0	Problematic/High Concern Area: Very Much - 2 Somewhat - 1 Not Much - 0	External Request For Information: Yes, Direct - 2 Yes, Indirect - 1 None - 0	TOTAL
Math Tutoring	1	2	2	0	5
Micro-Computer Reading Software	2	0	1	1	4
Student Selection Procedures	2	1	2	2	7
Teacher Diagnoses and Pre-scriptions	2	0	2	0	4

In this example, an ordered list of elements would be:

Student Selection Procedures (7)
Math Tutoring (5)
Microcomputer Reading Software and Teacher Diagnoses
and Prescriptions tied (4)

Student selection procedures is the highest ranked element and thus probably the most important to evaluate at this time. As resources permit, evaluation activities could be conducted on Math Tutoring, then the remaining elements, as indicated by their rank order.

Appendix 3-B

SAMPLE EVALUATION QUESTIONS

Purpose: Accountability and Reporting

1. Is the program in compliance with all of the funding agency's regulations and requirements?

2. How could reporting to funding agency be improved?

3. Are adequate records being kept of:

    demographic information about students?
    student participation?
    attendance?
    materials acquisition and use?
    amount of instruction?
    behavior/disciplinary problems?
    length of participation?
    student needs and student progress?

4. Has needs assessment data been used to plan programs?

5. Is the program serving the students which it was intended to serve?

6. On what basis have participants been selected?

7. Is staffing adequate to achieve program objectives?

8. Is data being collected to assess the impact of the program?

9. What are the relative costs of different program components?

10. How are costs related to program priorities?

11. What is the average cost

    per student?
    per hour of instruction?
    per unit of achievement?

12. How do costs vary by type of program?

Purpose: Assessing Short-Term Effects of Programs

1. What specific skills did students learn as a result of program participation?

2. How many skills do students master per unit time in the program?

3. Do students improve their performance on criterion-referenced tests as a result of participation in the program?

4. As a result of participation in the program, do students improve their performance on norm-referenced tests?

5. Do students change their attitudes about:

    the program?
    school?
    learning?
    self?
    control over their lives?
    reading?
    math?
    employment opportunities?
    future education?
    the institution?

6. Do students in the program learn significantly more than students who did not participate in the program?

7. Do students show changes in their classroom behavior?

    confidence?
    cooperativeness?
    study habits?
    interaction with teachers?
    independent study?
    time-on-task?

8. How do non-instructional services affect academic gains?

9. Do some types of students show greater gains than others?

Appendix 3-B (Continued)

Purpose: Assessing Long-Term Effects of Programs

1. Do some types of programs or methods produce longer term effects than others?

2. Do recidivism rates vary as a function of

    program participation?
    achievement level?
    attitude?
    amount and types of services received?
    other related issues?

3. Are program participants more successful at finding jobs? What kinds of students are most successful at finding jobs?

4. How do skills acquired in the program relate to the types of jobs students find?

5. Do a higher proportion of program participants return to school after release?

6. What kinds of students are most likely to return to school?

7. What skills do former program participants find themselves using most?

    life skills?
    job skills?
    academic skills?
    interpersonal skills?

8. In retrospect, how do former participants view strengths and weaknesses in the program?

9. What kinds of problems do former program participants face that might suggest changes in the program?

10. How do measures of cognitive gain vary across several years of program operation?

11. What changes have taken place in the program since its inception and what impact have these changes had on its effectiveness?

Appendix 3-B (Continued)

Purpose: **Determining the Degrees to Which a Program Has Been Implemented**

1. Does the program have formal written objectives?
2. Are program objectives realistic given constraints within the institution?
3. What are the key elements of the program and how many are in place?
4. In what different ways has each key element of the program been implemented?
5. To what extent do concerns about management routines (scheduling, location of materials, etc.) exist among staff?
6. To what extent do concerns about altering the program to better meet students' needs exist among staff?
7. Are classroom activities consistent with the objectives of the program?
8. What portion of instruction time is directly spent on tasks related to objectives of the program?
9. Do the materials being used match the objectives of the program?
10. What materials and activities have been most often used?
11. Is the mode of teacher-student interaction consistent with program objectives?
12. Is the program serving the students it was intended to serve?
13. Is information about the operation of the program being used to improve the program?

Appendix 3-B (Continued)

Purpose: Management and Staff Effectiveness

1. Are some teachers more effective than others with certain types of students? How can this differential effectiveness be used to improve instruction?
2. Are some teachers more effective than others in using certain types of materials or techniques?
3. In what ways is the administration of the institution being helpful or obstructive?
4. What kinds of in-service training programs have been or would be most useful to program staff?
5. How might improved management practices improve the quality of the program?
6. How might communication among staff members be improved?
7. How could roles and responsibilities be better defined?

Purpose: Determining How Best to Match Services with Individual Student's Needs

1. What are the predominant academic problems of students?
2. Does the nature of predominant problems vary by age of student?

   by socio-economic status?
   by program component?
3. What types of learning activities are effective for which types of problems? for students with different learning styles?
4. Do program activities accommodate a broad range of skill levels? Is there a sufficient variety of materials and learning activities for each skill level?

Appendix 3-B (Continued)

5. How well do students feel that the program is matching their needs? What improvements could be made?

6. Which component of the student needs assessment system is most useful to program personnel? Which components might be dropped or modified?

7. How effectively are individual student plans used to guide instruction (i.e., selection of materials and techniques)?

8. How well are non-cognitive needs being met?

Purpose: **Identification of Relationships Among Services and Program Components**

1. To what extent does this program supplement other programs?

2. In what ways could this program be better coordinated with other programs?

3. How do conflicts with other programs or activities affect participation in this program and how might they be overcome or minimized?

4. How does participation in this program affect performance or participation in other programs?

5. What factors influence a student's choice of services (given the option to choose)?

6. How much is information about students shared across programs?

7. What is the pattern of student referrals by one service component to others?

8. What services did students receive as a result of referrals from other program components?

9. What are service providers' perceptions of the effectiveness of other components? How accurate are those perceptions?

10. What are service providers' perceptions of the service/roles of other components? How accurate are those perceptions?

11. How might communication among programs be facilitated?

Purpose: Identification and Description of Effective Practices

1. How can data be collected to identify practices which lend themselves to

   positive attitude change?
   cognitive gains?
   behavioral change?
   long-term program effects?

2. Are some techniques viewed more positively than others by teachers? . . . by students? . . . by administrators?

3. Do techniques which are viewed most positively by students and/or teachers yield the best results?

4. How could staff make better use of effective practices (techniques and materials) from other institutions?

5. How could practices be disseminated to other institutions?

6. Are techniques differentially effective with different kinds of students (ages, sexes, types of offense, etc.)?

4. STRATEGIES AND TECHNIQUES FOR PROGRAM EVALUATION

INTRODUCTION

There are many strategies and techniques which can be used to evaluate a program, depending upon the purpose of the evaluation activities and based on the particular preferences held by those planning the evaluation. For just this reason and because there are no comprehensive evaluation models for Chapter 1 Neglected or Delinquent programs, this section is designed to present some basic approaches which can be used to evaluate selected program elements in order to provide information for reporting requirements and program improvement purposes.

It is recognized that interest in, and resources for, evaluation activities varies with the individual N or D program. Some programs are content to simply report descriptive data about their projects while others have attempted more ambitious evaluations designed to show the impact of the services and to determine where improvements might be made. Even among those N or D programs interested in conducting evaluations, there are basic philosophical differences about what strategies and techniques are appropriate given the unique conditions under which N or D services are often provided and the characteristics of the clients being served. Some argue that any evaluation of an N or D program has to be grounded in a standardized norm-referenced or criterion-referenced test that will yield student achievement data. Others feel that such test data is inappropriate for evaluating the N or D program and prefer to emphasize alternative approaches.

This section will not attempt to resolve the philosophical differences that exist over what evaluation strategies or techniques are best suited for N or D programs. Rather, it will present some approaches and techniques that can be used to collect information necessary to answer the evaluation questions of interest. The choice of approaches and techniques must be made in view of the evaluation questions being asked and is best left up to those responsible for the evaluation activities. It should also be noted that this section is a general overview of the selected strategies and techniques. Some are relatively straightforward and, depending upon the expertise of those conducting the evaluation activities, easily implemented. Others are more complex and might require further training or outside expertise in order to be implemented.

Assistance in planning N or D evaluations or training in any of these techniques can be obtained by calling a Regional Chapter 1 Technical Assistance Center.

TESTING APPROACHES

One fairly common approach used to evaluate the effect of a program is that of testing. In very general terms, some type of achievement test is administered in order to measure changes in student performance. If the improvement in performance is greater than would have been expected without the program, then the program is judged to have had a positive effect. So, tests might be used to collect information to answer the evaluation questions: "Do students in the program learn significantly more than students not in the program?" or "What specific skills did students learn as a result of participation in the program?"

This section will deal with the application of both norm-referenced and criterion-referenced testing approaches for program evaluation purposes. A variety of testing applications is presented, as well as possible constraints on using these approaches in the N or D setting. Information on specific tests is provided in Section 5.

Norm-Referenced Testing

A norm-referenced test is a test that is used to determine an individual's status with respect to the performance of other individuals on that test. This definition implies that the purpose of a norm-referenced test is to compare the performance of an individual with that of others. When standardized norm-referenced tests are used in a school setting, this comparison is usually made with those of a similar age or grade level. The performance of the comparison group is found in a table of norms which have been derived through previous administrations of the test to selected school samples.

Problems in Using Norm-Referenced Tests in N or D Evaluations. There are a variety of potential problems associated with using norm-referenced tests in N or D settings. Before making decisions regarding the use of a norm-referenced test as part of the evaluation activities, these problems should be considered.

71

1. The published norms are usually inappropriate for use with an N or D population. The group from which the norms are derived generally represents an average school population. Since an N or D group is atypical and may not perform like a typical school group, comparisons made with such a group may be difficult to interpret.

2. In order for test scores to be comparable with the norms, the test should be administered within specified testing periods during the year. As students are continually entering and leaving the N or D setting and staying for varying periods of time, it is often impractical to restrict the testing to these specified dates.

3. The test is supposed to be administered in a precisely defined way. The mode of operation of an N or D institution and the characteristics of its students often make it very difficult to follow the standardized testing procedures.

4. The N or D students may not always be motivated to do their best on a norm-referenced test. In addition, for a variety of reasons, students will often exit from the program without receiving a posttest.

5. Norm-referenced tests are generally meant to serve many different programs and a variety of populations; therefore their content may be too broad to adequately test the specific skills covered in the N or D program. As a result, the amount of real improvement may be underestimated by these tests. A norm-referenced test will be especially insensitive to the small improvements made by the many N or D students who are in the program for only a brief period of time.

6. The items on a norm-referenced test are typically selected to spread out the range of scores so that individual-to-group comparisons will be facilitated. This wide range of items often makes it difficult to relate a student's score to specific instructional needs.

7. The language used in norm-referenced tests is not always appropriate for N or D students.

> The skill deficiencies of the N or D student will sometimes require the administration of a test that is considerably below the student's grade level. Here the student encounters language and situations aimed at much younger students, and resentment may develop. Some tests have now appeared which attempt to overcome this problem by using a high-interest low-ability approach.

Ways to Use Norm-Referenced Tests in N or D Evaluations. Four ways of using norm-referenced tests are described in this section. The first concerns the determination of current status using the norms tables, while the other three pertain to the determination of overall program effect. Any method chosen should be appropriate to the particular program being evaluated and should provide information bearing on the evaluation questions at issue. In many cases it may be apparent that the use of criterion-referenced tests, to be described later, will yield more useful information.

1. The Use of Norm-Referenced Tests to Determine Current Status -- Sometimes it may be of interest to determine how students' performances in an N or D program compare with those of a typical school population. The types of evaluation questions being asked might include: "How far behind their public school peers is this group of N or D students?" or "In which subject areas do the N or D students need special work?" For either question, norm-referenced testing could be appropriate. Similarly, if the evaluation activities are focusing on the element of student selection and the question is: "Which students are most in need of participation in the program?", norm-referenced testing could again be appropriate.

   When using norms to determine current status it is important to test the students within the dates for which the norms are established and to adhere as closely as possible to the standardized testing procedures. It must be remembered that the comparison is being made with a different population of students and any interpretation of the results should take this into account.

2. The Use of Norm-Referenced Tests to Determine Overall Effect -- Here the N or D students are given both a pretest and posttest. The position of these students relative to the norming group is determined both at pretest time and at posttest time. Any improvement in their position is assumed to be due to the special educational treatment being provided in the program. Generally, the types of evaluation questions being asked might include: "Do students in the program learn significantly more than students who did not participate in the program?" or "How do measures of cognitive gain vary across several years of program operation?"

Both the pretest and the posttest must be administered within the testing periods for which norms are established, regardless of when students enter or leave the program. Norming periods generally occur near the beginning or toward the end of the school year and, for some tests, the middle of the year as well. Thus students would have to be in the program for several months in order to receive both the pretest and posttest. Here again, the standardized testing procedures should be followed as closely as possible and interpretations should take into account the fact that comparisons are being made with a different population of students.

3. **The Use of Norm-Referenced Raw Scores and Standard Scores** -- It is possible to obtain a measure of overall program effect without referring to the published norms. Here again, the N or D students are given both a pretest and a posttest, but it is the change in their raw scores or standard scores that is used to determine program effect. Where a choice exists, the standard scores should be used since their statistical properties are more amenable to the calculation of score changes. The types of evaluation questions being asked might include: "Do students improve their performances on the test as a result of participation in the program?" or "Do students need more of an instructional emphasis on a particular subject matter?"

   As adherence to specified testing periods is not required with this approach, students can be pretested and posttested upon entry into and exit from the program. In addition, students with only brief stays in the program can now be included in the analysis.

   The major problem with this approach is the difficulty of interpreting the meaning of a particular gain, whether in raw score or standard score units, without reference to norming information. To say that an N or D group has gained nine standard score points does not convey a great deal without knowing what others have done. One way of dealing with this problem has been to convert raw scores to grade equivalent scores and to express gains in terms of grade equivalents. This approach is definitely not recommended due to the misinterpretations associated with the use of grade-equivalent scores.

4. **The Use of the Systematic Allocation Model** -- This model can be used to evaluate an educational program within an N or D setting if there exists within the setting students who are not in the program and if students are selected for the program on the basis of need. The type of evaluation questions asked might include: "Do students in the program demonstrate a greater level of academic improvement than those not in the program?"

It should be noted that the Systematic Allocation Model can also be used with criterion-referenced tests. Application of this model with a criterion-referenced test eliminates the problems associated with norm-referenced testing and the N or D program. (For a detailed discussion of the Systematic Allocation Model see the Handbook for Evaluation of Title I Programs in State Institutions for Neglected or Delinquent Youth, 1978. This Handbook was developed by the System Development Corporation, under U.S.O.E. contract number 300-76-0093.)

Criterion-Referenced Testing

A criterion-referenced test is a test that is used to determine an individual's status with respect to specified objectives of instruction. For each objective assessed in the test, a set of items is developed to determine whether the student has, in fact, mastered that objective. Prior to the administration of the test, a criterion is established which is then used as the standard to determine whether the resulting scores indicate mastery or non-mastery of each objective tested. A criterion-referenced test may cover one or more instructional objectives, depending upon the purpose of the test.

In deciding whether to use norm-referenced or criterion-referenced tests for evaluation purposes, the types of information that each will provide should be taken into consideration. With criterion-referenced testing, each student's score is compared to the prespecified standard to determine mastery, while in norm-referenced testing the student's score is compared to that of the norming group. This means that in criterion-referenced testing the resulting scores indicate which instructional objectives have been mastered, while in norm-referenced testing the resulting scores indicate the student's position relative to the scores of the norming group.

Typically, in norm-referenced testing the focus is on individual-to-group comparisons, test content is general and may not be matched to a particular instructional content, and items are selected to deliberately spread out the score distributions. With criterion-referenced testing, the focus is on individual comparisons to prespecified standards, the test content is much more specific and thus easier to relate to a particular instructional content, and the items are not selected to spread out the range of student scores.

Uses of Criterion-Referenced Testing for Program Purposes. There are many ways to use criterion-referenced tests in order to collect information useful both for student assessment and for the evaluation of the N or D program. Four general applications are described briefly and should be kept in mind when making testing decisions.

1. Criterion-referenced tests can be used to assess the strengths and weaknesses of individual students as they enter the educational program. Information regarding the degree to which various instructional objectives have already been mastered will help in developing the student's individual plan. This same information can also be used to plan and evaluate the instructional programs.

2. Criterion-referenced tests can be used to assess the status of individual students as they complete segments of the instructional program. This information can then be used to determine whether additional instruction is required in that area or whether the student is ready to move on to a new area of instruction. This same information can also be used to evaluate the effectiveness of the instructional program and the curriculum materials being used within the program.

3. Criterion-referenced tests can be used to assess the strengths and weaknesses of a group in order to determine where to place emphasis within the overall instructional program. This type of information can be used in program development, grouping, staff assignments, and in the evaluation of the later success of the instructional interventions.

4. Criterion-referenced tests can be used to determine the overall effect of an educational program. The information obtained can be used to describe program effects and as a planning guide for possible program improvements. The processes involved in this type of evaluation are discussed in the following section.

Ways to Use Criterion-Referenced Tests in Evaluating Overall Program Effect. Two ways of using criterion-referenced tests for evaluating program effects are discussed in this section. Both methods have the advantage of not being restricted by the problems usually associated with norm-referenced testing, and both will provide relevant information for evaluation purposes.

1. **The Use of the Systematic Allocation Model** -- As discussed earlier, this model can be used to evaluate an educational program within an N or D setting if there are students within that setting who are not in the program and if students are selected for the program on the basis of predetermined index of need. The types of evaluation questions that might be asked include: "What specific skills did students learn as a result of program participation?" or "Do students in the program learn significantly more than students who did not participate in the program?"

   Application of this model has already been described in the section on norm-referenced tests. The model is applied in exactly the same way when criterion-referenced tests are used. However, the use of criterion-referenced tests has the advantage of allowing for the simultaneous application of the Criterion Model, as described next.

2. **The Use of the Criterion Model** -- This model can be used to evaluate an educational program within an N or D setting when the type of comparison group required in the Systematic Allocation Model is not available. This would be the case either when all or practically all students are assigned to the program, or when assignment to the program cannot be based on the cutoff score used as the index of need. The types of evaluation questions that might be asked include: "How many basic reading skills do the students master during the first month of the program?" or "Which instructional objectives are still not mastered by students after leaving the program?"

   The Criterion Model requires that a performance standard be set in advance for the criterion-referenced test being administered to the students in the program. That is to say, what the group is expected to accomplish in terms of mastery on the test must be stated in advance. Criteria may be established in a variety of ways, such as based on prior performance of a similar group or by teacher judgment as to what should be expected. Indirectly then, expectations regarding the group's level of mastery on the objectives being tested must be set in advance. For example, in a basic skills mathematics program, and with reference to a particular criterion-referenced test, it may be decided that 80% of the students in the program should be able to score 75% correct on the addition and subtraction items and 70% correct on the multiplication and division items. Students are pretested with the criterion-referenced test as they enter the program and posttested with the same test as they leave the program. The posttest results are compared with the pre-established standard to determine which of the criteria have or have not been met.

Pretest and posttest results are compared to determine what kinds of improvements have taken place in the students' levels of performance. Information thus collected can provide input toward program change and improvement.

It should be noted that the Criterion Model does not provide a definitive evaluation of the educational program, since no comparison group is involved. The results observed could be due to other instruction received or to outside causes. Nevertheless, the information provided by application of the model is suggestive of program effect and surely provides direction for program evaluation and improvement.

N or D programs which are able to use the Systematic Allocation Model can simultaneously apply the Criterion Model and thus obtain the additional information this model provides. In order for this to be done, of course, a criterion-referenced test must be used when implementing the Systematic Allocation Model. (For a more detailed discussion of either model see the _Handbook for Evaluation of Title I Programs in State Institutions for Neglected or Delinquent Youth_, 1978.)

Constructing a Criterion-Referenced Test. Sometimes it is necessary to construct a criterion-referenced test for the purposes of the N or D evaluation. In general, this is recommended only when evaluating the effect of a short unit of instruction. When evaluating the overall effect of an instructional program it is better, if possible, to use a published criterion-referenced test. This is because the construction of a statistically sound criterion-referenced test which is appropriate for evaluating an entire program is an involved proposition which requires much in the way of time and resources. It therefore is advisable to initially review the available published tests to determine whether they meet the evaluation needs.

In developing criterion-referenced tests for use in evaluating individual units of instruction, the following steps are recommended. The steps are only briefly described here in order to provide an overview of the process involved.

1. _Select the objectives to be measured_. These objectives should be ones taken from the instructional program being implemented or to be implemented. The objectives should be stated in measurable terms, clearly indicating the expected learning outcome of the student.

2. Develop the test specifications. Specify what the test is going to look like. Determine how many items will be included for each objective, the item format(s) to be used (e.g., multiple choice, true-false), the reading level, vocabulary, organization of the items, and how the student will respond (e.g., write the letter of the correct answer, circle the correct answer).

3. Develop the items for the test. At this point the items are constructed in accordance with the test specifications. Each item should assess some aspect of the objective(s) being measured. Sometimes it is possible to obtain previously developed items from item banks which have been established for just this purpose. (A brief discussion of item banks is included later in this section.)

4. Check the appropriateness of the items. The content of the test items should be reviewed by other instructors who are familiar with the subject in order to determine whether they are appropriate in content, vocabulary and format. It should also be determined whether each objective has been adequately tested by including enough items. The items should also be tried out with a small sample of students to determine whether there are any problems in interpretation. On the basis of results of instructor and student reviews, the items should be revised as necessary.

5. Assemble the items into a test. Decide on the test layout and put together the actual test. Ensure that appropriate directions have been included for each section of the test, prepare any necessary answer sheets, and develop a scoring key.

6. Establish the standards for interpreting the test results. Determine, in advance, what will be expected of each student in order to be classified as having reached mastery. If the test includes more than one objective, decide (for each objective) how many items must be answered correctly in order to say that the objective has been mastered. If the test covers only one objective, it will be satisfactory to indicate a percentage or total number correct as the criterion for mastery on the test.

Item Banks and Their Use for Constructing Tests

Simply stated, item banks are collections of test items which have been developed to assess the mastery of specified objectives. These item banks may be developed and maintained by commercial publishers, in which case the items will typically be sold to interested users. The item banks may also be maintained by a non-profit group, in which case the items are available for a minimal fee or on a trade basis; donating items to the bank allows withdrawal of others. In any item bank, the items are usually grouped by subject·matter and the specific objective being tested.

In general the idea of using an item bank to develop a test is a sound one. If the number of items written yearly by instructors to assess student achievement were to be counted, the total would probably include thousands of items. If possible, rather than writing new items each time a test is constructed, it would be a better use of time to go to a bank to select items which have already been developed; hence, the introduction of the item bank.

To construct a test using an item bank, the instructor follows the same general steps as described in the section on criterion-referenced test development. First, the instructor must identify the objectives to be measured. Then the test specifications are developed. However, instead of next developing the actual items, it is here that the instructor makes use of the item bank. With an item bank, the instructor chooses items that have already been written and that match the test specifications. This should save considerable time and effort on the part of the instructor. Once the items have been selected, the appropriateness of the items should be determined, the items assembled into the actual test, and the standards for interpreting the test results established.

Deciding Whether to Use an Item Bank. In considering whether to use item banks for local test development purposes, one point should be in mind: the items will vary in quality from item bank to item bank. Not all item banks screen items to eliminate those of poor quality. In fact, some item banks accept any items without screening or editing. So, while the use of item banks can save much in the way of time and effort and often results in high quality tests, there are certain questions that should be asked before using a particular bank.

80

Appendix 4-A includes a general list of questions that should be asked in determining whether existing item banks would be useful. One item bank may not meet all the requirements. As with any standardized test or set of instructional materials designed for general use, the instructor will need to identify the item bank that best suits the present program needs. For a brief description of sources of item banks see Section 5.

ALTERNATIVE TECHNIQUES

Although tests are useful tools for evaluating N or D programs, there are often occasions when test data alone will not be sufficient to answer the evaluation questions of interest. This will be especially true if the evaluation activities focus on the understanding of a process, attitudes of those involved, or other types of behaviors that do not lend themselves to being measured through achievement type tests.

Consider, for example, the evaluation question: "Does the program contribute significantly to improvement in reading skills?" This question could be answered through some type of testing activity. However, if the question were "Does the program contribute significantly to improvement in reading skills and positive attitudes toward reading?", test data alone would not be sufficient. To collect information relating to the question of whether student attitudes toward reading have changed, other types of information need to be collected. This might include a systematic examination of library records to determine whether more books have been checked out by the students in the last month, or a series of interviews with the students in the program.

So, often it will be necessary to collect something other than test information in order to fully answer the evaluation questions of interest. This section will briefly discuss four techniques which can be used to collect other evaluation information, including: observations, interviews, questionnaires and existing records. The guide to selecting the most appropriate technique will be the evaluation questions being asked. Additionally, because of its great relevance to N or D programs, the concept of time-on-task is discussed as a special application of observations.

Observations

The observation is a method for collecting information by systematically watching what is occurring at certain times. The patterns of behavior being observed may range from the very simple (such as recording whether the student is in the appropriate place) to the very complex (such as classifying exactly how two students are interacting). The person doing the observations may be the instructor, a volunteer aide, a parent, an evaluator, or even another student. In the context of the N or D program, the use of observations helps focus the data collection activities on areas not so easily measured by tests -- areas such as: student enthusiasm toward certain instructional approaches, the quality of interactions between the students and the instructor, and, the amount of time spent on instructional tasks. The use of observational data can contribute toward a better understanding of why or how something happens and can also document that the event did occur.

When considering whether to use observations to collect evaluation information, the following four points should be kept in mind:

- Observations provide a means of collecting information that would not be available through other techniques (e.g., the number of times positive verbal reinforcement is used with students).

- Observations provide a means of collecting information which does not rely on recall of what might have happened in the past, reducing the chance that events may be forgotten, overlooked, or distorted over time.

- Observations provide a means of collecting information that removes individual points of view from the data (e.g., having an observer record and classify the types of instructor-student interactions rather than asking the instructor about the types of interactions provides more objective data).

- Observations provide a means of collecting information in a variety of settings and with many types of individuals, where other data collection techniques may not be appropriate (e.g., while it might be difficult to successfully interview a group of students in a work setting, an observer could document their behavior through observations).

While there are clearly advantages to using observations to collect information, there are also some constraints involved in this approach. When considering whether to use observations to collect data, the following points should be considered:

- The actual presence of an observer may alter the behavior of those being watched.
- The observer may interpret behavior in a way that is different from those actually involved in the activity.
- The observer needs to be trained in how to observe and record behavior; the more complex the observation system being used, the more time-consuming this training may be.
- When fairly complex behavior patterns are being observed, the reliability of observers can be an issue in interpreting results.
- When the presence of the observer is required for long time periods, this technique may be an expensive way to collect results.
- Because observations require a significant time commitment, the sample size used may have to be smaller than that used with other techniques.

If the evaluation questions of concern seem best answered by watching for certain events, observations should be used to collect the information. Appendix 4-B includes some very brief guidelines for developing and conducting observations.

Time-on-Task

One major instructional factor relating to achievement is the amount of time a student spends actually engaged in tasks which further his or her skills. Time-on-task, then, is the time devoted to tasks directly related to the development of the desired skills. Student performance can be improved by increasing the time spent on actively learning and practicing a skill. In general, as reflected in higher test scores, students learn more when they spend more time engaged in learning activities.

As wouuld be expected, the amount of time students spend in learning differs dramatically from classroom to classroom. While a student may be scheduled to attend a class for a certain period of time, a variety of activities other than learning may occur during that time. For example, students may be engaged in socializing, obtaining materials with which to work, recordkeeping, being disciplined, or other non-instructional activities. Of course all class time can not be used directly in learning skills, but a good portion of time should go toward skill acquisition.

When evaluating a program it may be necessary to determine how time is actually being used in the classroom. If, for example, the evaluation question is: "What portion of instructional time is directly spent on tasks related to objectives of the program?" or "How effective are the instructors in managing classroom activities?" then it would be necessary to determine how time is actually used. Likewise, if the results of past evaluation activities have indicated that students are not improving their skills, the use of class time may become an issue. In any of these situations, in order to answer the questions being asked, it will be necessary to systematically observe the classrooms in order to document what is occurring. Clearly the purpose in determining how time is spent in the classroom is to increase the amount of time-on-task, thus increasing student learning. To accomplish this it is necessary first to determine exactly how time is being used and then to reduce the non-instructional uses where possible. So, when observing the classroom it will be necessary to document how much time is devoted to learning and how much time is used for other activities.

Engaged time can be used on interactive activities in which the student is working on instructional tasks with others (e.g., the instructor, an aide, or other students) or on non-interactive tasks in which the student is working alone. Engaged time, then, includes activities such as:

- competing in drill and practice games,
- participating in a discussion,
- taking part in role-playing activities,
- listening to a lecture,
- asking questions,
- participating in a demonstration,
- receiving feedback on some work,
- reading an instructional manual,
- working on written assignments,

- watching a filmstrip, or
- working with a microcomputer program.

Non-engaged time, on the other hand, includes activities such as:

- working on other assignments,
- socializing with others,
- obtaining work materials,
- answering to roll call,
- filling out health forms,
- being disciplined,
- recording progress,
- passing out papers,
- being called out of the room,
- observing others, or
- doing nothing.

While some of these activities, such as obtaining work materials, are necessary to the task of learning, others are not a very effective use of class time and should be reduced whenever possible.

After the time-on-task observations have been completed it will be possible to look at the total picture of student activities to determine how to increase the effective use of class time. How much time is spent on instruction? On non-instructional tasks? Do students take too long to get ready to work? Are too many administrative activities reducing the time available for instructional purposes? Understanding how time is really used will indicate where changes could be made in order to increase the amount of time students spend engaged in instruction and, ultimately, will increase student learning.

Questionnaires

The questionnaire presents individuals with a series of carefully developed questions covering a predefined topic and requires some manner of written response. Questionnaires may include items that are open-ended (requiring the respondent to write in some form of response), closed (requiring only that the respondent select an answer from choices provided), or both.

For N or D programs, the use of questionnaires allows the collection of information through methods such as: surveying those who employ graduates of the program in order to determine employer satisfaction and to pinpoint areas in which changes might be made to improve the program; surveying past students to determine whether the program was effective for them; determining community attitudes toward the program; and documenting parental attitudes toward the program. In general, the use of questionnaires facilitates the collection of a wide variety of information from a large group of individuals.

In considering whether to use questionnaires to collect evaluation information, the following five points are of importance:

- Questionnaires provide an inexpensive means of simultaneously collecting information from a large number of people.
- Questionnaires provide a means of ensuring the respondents' anonymity, which sometimes results in more honest responses to sensitive questions.
- Questionnaires provide a means of asking uniform questions to everyone, thus ensuring that the necessary data is collected from all involved.
- Questionnaires, especially those using closed item formats, provide a means of collecting data which is fairly easy to summarize and interpret.
- Questionnaires provide a means of collecting information over many topics of interest, ranging from general attitudes to details on past experiences.

Clearly there are many advantages to using questionnaires, but there are also constraints which should be noted, including:

- If a question is unclear to the respondent, it cannot be clarified and as a result might go unanswered.

- Those asking the questions may not be able to follow up on interesting lines of thought or probe for more detailed responses.

- Because questionnaires are somewhat impersonal, the response rate may be low and those who do respond may constitute a biased sample.

- Some respondents may have difficulty in reading or in expressing answers in writing.

If the evaluation questions of concern seem best answered by asking a number of people a series of written questions, questionnaires should be used to collect the information. Appendix 4-C includes some very brief guidelines for developing and administering questionnaires.

Interviews

The interview is a method for collecting information by asking a series of questions of each individual included in the sample. Rather than requiring the respondent to read a question and answer it in writing, the interviewer asks each question, carefully recording the oral response given. The interviewer may then systematically follow up on the responses, either through an informal approach or through a predetermined set of additional questions.

For N or D programs, the interview facilitates the collection of information from individuals who may have difficulty reading or writing, where non-verbal reactions are highly relevant, and where detailed probing of responses is necessary. Interviews would be an appropriate means of collecting information on areas such as: the types of instructional approaches that the students prefer; the types of interpersonal skills that potential employers would like future employees to have; or the ways in which instructors would like to see the program reorganized.

When considering whether to use interviews to collect evaluation information, the following six points are relevant:

- Interviews provide a means of collecting information which does not depend upon the reading or writing skills of the respondent.

- Interviews provide a means of collecting information in a manner which first allows rapport to be established between those involved.

- Interviews provide a means of collecting information on non-verbal responses, language and voice inflection, as well as the verbal response.

- Interviews provide a means of collecting detailed information through the use of rephrasing of questions and further probing of responses.

- Interviews provide a means of asking questions that may be difficult to phrase in writing or which require fairly extensive clarification.

- Interviews provide a means of collecting all of the information from those who participate, avoiding the possibility that responses may be missed due to unclear questions.

Interviews, of course, have constraints which should be considered before deciding to use the techniques. Some of the constraints are as follows:

- Because the results of the interview rely strongly on the interviewer's interpersonal skills and communication capabilities, some respondents may be threatened, led toward certain responses, or be generally uncommunicative.

- The interviewer, unless carefully trained, may get off track, alter the meaning of questions by slightly rephrasing a few words, fail to follow up on responses where appropriate, or miss key non-verbal nuances.

- Because this technique requires the presence of an interviewer at all times, the interview can be an expensive and time-consuming way to collect information; smaller sample sizes may be necessitated.

- Without careful planning, interview data can be difficult to summarize and interpret.

If the evaluation questions being asked seem best answered through the oral administration of a set of questions and further probing on responses, then interviews should be used to collect the information. Appendix 4-D includes some very brief guidelines for developing and conducting interviews.

Existing Records

Existing records, although not an information collection technique, are clearly a source of evaluation information. The use of existing records is a method of collecting information on the basis of what has already been compiled in some manner by others. This would include any kind of data which has been systematically collected at a previous date or obtained as a byproduct of other activities.

Existing records encompass a broad range of information, including: records maintained for student management (e.g., class or individual progress charts, individual student files, student scores on progress tests); records maintained for short- and long-range planning purposes (e.g., objectives taught in the program, services available outside of the program, staff backgrounds); and records maintained for evaluation and administrative reporting purposes (e.g., attendance, discipline reports, standardized achievement tests, funds used for specialized equipment).

In the N or D program, existing records can provide a source of information to answer many types of evaluation questions. (Section 6 deals in detail with the types of records which should be maintained and their use for program evaluation and management.) Further, existing records can provide the background information necessary to complement data collected in other ways. For example, while interviews may be used to collect information on students' attitudes toward certain instructional materials, existing class records can document how well the students have learned by using the various materials. Combined, the two pieces of information provide a more complete picture for evaluation purposes.

When considering whether to use existing records for evaluation purposes, the following four points should be kept in mind:

- Existing records provide a wide variety of readily available information.
- Existing records, because they have not been interpreted by others, are generally a source of fairly objective information.
- Existing records are generally considered a credible source of information because the data has been collected at the time of the event, rather than recalled at a later date.

- Existing records provide information which is low in cost to collect and may be obtained in a shorter time period than that required to get new information.

There are, of course, constraints which should be kept in mind when considering the use of existing records, including:

- Existing records may be incomplete, with generally no way to retrieve this missing information.
- It may take some time and effort to extract the desired information from the existing records (e.g., the last five years of the test scores may be available, but stored in a box in the basement of another building).
- Permission to use existing records may involve some legal requirements, such as permission from the individuals whose records are of interest.

For all evaluation questions being asked, consider the feasibility of using existing records either to answer the question or to provide supplementary information necessary to fill out the picture. Appendix 4-E includes some very brief guidelines for the use of existing records.

Appendix 4-A

QUESTIONS TO ASK REGARDING AN ITEM BANK

1. What kind of information is available about the individual items in the bank?

   The information needed about items will differ depending on the academic area. Constructing math tests from item banks, for instance, does not require as much information about items as constructing reading tests.

   a. Can the appropriate grade level be identified for which an item is appropriate?

      This is usually not difficult for math tests; math items are usually described by a particular operation that is taught at a certain grade level. For reading items, however, if the instructor is looking for an item where the student must identify the main idea, he or she will probably want a way of knowing the reading level of a passage without having to actually pull the item from the bank first.

   b. What kind of information is available about the technical quality of the items?

      Can the instructor tell how difficult the item is for different grade levels of students? Most instructors prefer that the test contain both easy and difficult items to allow students to show what they do know and to find what they do not know. Also, a check must be made to see if this information about the item has been updated.

   c. Is there a way that instructors can use the students' incorrect answers to diagnose their problems?

      Often instructors like to use the results from a test in the diagnosis of their students' strengths and weaknesses. It helps if the instructor can identify a problem by using information from a wrong answer that was chosen in the test.

d. How specific can the instructor be when requesting items?

For example, suppose the instructor would like an item to measure the recognition of the consonant blend "CL". Is it possible to pull items directly from the bank that deal with "CL" or must all items dealing with consonant blends be searched through to locate items dealing with "CL"?

2. What kinds of quality control measures have been applied to the items entered into the bank?

Have the items been reviewed by instructors and curriculum experts for correctness of the answers? Is there a guarantee that the items really do measure the skills they profess to measure? Have the items been reviewed for possible biases such as toward different sex, ethnic, racial or regional groups?

3. What kinds of item response formats are available?

Does the bank include items in a variety of response formats and is there an option when choosing the items?

4. What are the actual procedures that must be followed when using the item bank?

Some developers of item banks request the requirements for a test and will deliver either the options for the items or the actual test. Others will supply the actual item bank.

A good suggestion here is that when an item bank is considered, a test run should be made involving the actual persons who will be using it. Records should be kept of what has to be done, how long it takes, how difficult it is and how it compares to what has been done in the past.

5. How does the organization of the item bank match the instructor's curricular organization?

How difficult will it be to locate the sections of items in the bank that deal with particular sections of instruction? In some cases, the instructor will find that it is easier to adopt the objective system of the item bank than to translate program objectives to the objectives of the item bank.

Appendix 4-A (Continued)

6. What is the cost?

   Developing an item bank can be a very costly venture. Buying an item bank or contracting with an item bank service can also be expensive. A careful analysis should be made to determine whether the advantages of the item bank outweigh these costs.

Appendix 4-B

GUIDELINES FOR DEVELOPING AND CONDUCTING OBSERVATIONS

A. WHEN PLANNING THE OBSERVATION:

1. Identify the category of behaviors on which data will be collected. Limit the category to one small enough to be reasonably done during an observation session. Do not expect to collect information on every behavior of interest at one time.

2. Determine who will be observed. The sample will affect how the observations are done, the length of the observation, and the system for doing the actual observation.

3. Decide ahead of time how the results of the observations will be analyzed. The data analysis can affect the format of the observer recording sheets and the types of information actually collected.

4. Limit observations to areas in which information cannot be collected in other ways. For example, using observations to obtain information on the age of some students would not be the best method of data collection. On the other hand, direct observation to determine eye-hand coordination of students would be appropriate.

B. WHEN DEVELOPING THE OBSERVATION INSTRUMENT:

1. Identify and clearly define each behavior that the observers will be looking for. The explanation of a behavior should not be vague, general, or open to interpretation by the observers. Having observers watch for disruptive behavior would result in very unreliable data. Exactly what is disruptive behavior? In comparison, having observers tally the number of times a student left the seat would be a behavior much less open to interpretation, resulting in more reliable data.

2. Develop a coding method, tally sheet, or other device that facilitates the observation process. If the observer has to take time out to write down words, the behaviors occurring during that time will be lost.

C. WHEN PLANNING TO USE THE OBSERVATION:

1. Ensure that each observer is fully trained in the procedure. This would include an understanding of the definitions of each behavior, practice at using the device on which the data will be recorded, and how to be unobtrusive while doing an observation. If the purpose of the observation is simply to describe events as they occur rather than watch for specific behaviors, the observer still needs to be trained in methods for recording behaviors.

2. When scheduling the observations, keep each period fairly short. Observing and recording behaviors is a very intense activity, so should be divided into several brief periods, rather than one long one. For example, if the observer needs to watch a classroom for a total of 30 minutes, ten three-minute series of observations would provide better data than three ten-minute periods. Of course, there may be times when the purpose of the observation is to describe what went on during an entire lesson, in which case it would not be possible to break up the observation periods.

Appendix 4-C

GUIDELINES FOR DEVELOPING AND ADMINISTERING QUESTIONNAIRES

A. WHEN PLANNING THE QUESTIONNAIRE:

1. Identify the topic of the questionnaire. Decide on this topic before beginning to develop the questions and stay within that area.
2. Determine the intended audience before the questions are developed. Audience characteristics will affect the format of the entire questionnaire and the phrasing of each question.
3. Determine ahead of time how the questionnaire results will be analyzed. This will affect the format of the included questions. For example, if the questionnaire responses are going to be machine-scored, open-ended questions could not be used. Or the closed format might be used and respondents asked to answer on a separate sheet which could then be scored directly by the computer. The questionnaire should also be arranged to facilitate scoring responses by grouping similar items together (i.e., all yes/no type questions together).
4. Ask only for information which cannot be obtained elsewhere. The purpose of the questionnaire is to collect some type of information or attitudes from each individual. Each person is responding because his or her input is necessary and of interest. If the information can be obtained elsewhere there is no reason to have a person spend time repeating that information.
5. Keep the questionnaire short. A person is much less likely to respond to a long questionnaire and more likely to return the questionnaire if it is of reasonable length.

B. WHEN DEVELOPING THE QUESTIONS TO INCLUDE ON THE QUESTIONNAIRE:

Order the questions in a logical manner. Start with the most general types of questions, then move on to the specifics.

2. Limit each question to one idea. Do not combine more than one idea in a single question. If a question does cover more than one point it is impossible to interpret the results later.

3. Do not ask leading questions. Be sure that the questions do not lead the respondent toward the desired response. If the respondent can tell what the "correct" response is, then the question should be rewritten.

4. Word each question as simply and clearly as possible. Do not include information that is unnecessary to the question and avoid technical terms, unless they are appropriate to the audience.

5. Include a definite point of reference to ensure that each individual responds to the same question. For example, if a question is asked, "How many hours do you work?," respondents may answer in terms of hours per day, per week, or in other ways. Changing the question to ask "How many hours per day do you work?" ensures that each individual responds to the question in the same manner.

6. If a closed question format is used, try to include options that cover all possible aspects of that question. Do not limit the answers to only one side or part of an issue. Additionally, since it is often difficult to anticipate all possible choices to include in a closed format, use the category "other" and allow a space for the person to write in a response.

C. WHEN PLANNING FOR ADMINISTRATION OF THE QUESTIONAIRE:

1. Include complete and clear directions on how to respond to the questionnaire. Explain exactly how to respond and where. Do not leave anything up to the respondent's imagination.

2. Include a cover letter with the questionnaire. This letter should be addressed to each respondent. The purpose of this letter is to establish rapport with the respondent, to explain why the questionnaire is being sent, and to

encourage that the questionnaire be returned. It is a good idea to include a deadline date for returns in this letter.

3. Ensure that the questionnaire and all other correspondence is neat and easy to read. A poorly arranged questionnaire or one that is difficult to read will have less of a chance of being returned than one which is well-designed.

4. Include a stamped, self-addressed envelope with the questionnaire. This will help encourage returns.

5. Use postcards or other means to follow up on those questionnaires that were not returned. Remember that the more questionnaires returned, the less biased the sample.

Appendix 4-D

GUIDELINES FOR DEVELOPING AND CONDUCTING INTERVIEWS

A. WHEN PLANNING THE INTERVIEW:

1. Identify the one topic of the interview.
2. Determine the intended audience.
3. Determine how the results will be analyzed.
4. Ask only for information which cannot be obtained elsewhere.

B. WHEN DEVELOPING THE QUESTIONS TO ASK DURING THE INTERVIEW:

1. Order the questions in a logical manner.
2. Limit each question to one idea.
3. Do not ask leading questions.
4. Word each question as simply and as clearly as possible.
5. Include a definite point of reference to ensure that individuals respond to the same questions.

C. WHEN PLANNING FOR ADMINISTRATION OF THE INTERVIEW:

1. Establish a method for recording interviewee responses. It is important that the actual wording be preserved as closely as possible. Recording methods include taking notes during or after the interview, using preplanned recording sheets, or taping the session.
2. Train the individuals conducting the interview. They should be able to conduct each interview in the same manner, ask the same questions in the same order, and avoid any emotional responses to the interviewee's answers which might affect future responses. The interviewer must be trained to constantly probe for additional information and accurately note responses to each question.

The interviewer should also be trained to use the first few minutes of the interview to establish a good rapport with the interviewee by explaining the purpose of the interview.

Appendix 4-E

GUIDELINES FOR USING EXISTING RECORDS

A. WHEN DECIDING TO USE EXISTING RECORDS:

1. Identify the area of information for the data collection. Once the area is established, appropriate types of existing records can be selected.

2. Describe the sample on which the information will be collected as clearly and completely as possible (for example, fifth grade students who have attended elementary schools in the district for the last two years).

3. Decide exactly what type(s) of records will be used. Considering feasibility, cost, access, time, and legal implications will help in determining this.

4. Determine ahead of time how the information will be analyzed. The process used for data analysis will have implications for how the information is recorded.

B. WHEN PLANNING TO COLLECT INFORMATION FROM EXISTING RECORDS:

1. Identify exactly where the necessary information is located and whose permission must be obtained in order to access these records.

2. Determine ahead of time any legal requirements which must be met in collecting or using the records.

3. Develop a method for extracting the necessary information and a means of recording that data. The method used must be easily understandable and consistent.

4. Train those who will be collecting the information. To obtain reliable data, each individual must record the same type of information. Therefore, each person must clearly understand any categories on the recording forms and directions for their use.

5

INSTRUMENTATION

5. INSTRUMENTATION

INTRODUCTION

Selecting the appropriate type of instrumentation is a very important part of the evaluation process. Section 4 discussed the application of testing and other types of information collection techniques used to answer evaluation questions. This section provides specific information on criterion- and norm-referenced tests, measures of affective behavior and sources of item banks. Selected references for further information on instrumentation are also provided.

AN ANNOTATED BIBLIOGRAPHY

The bibliography which follows presents information about the various tests and other measurement techniques available for evaluation of N or D programs.

> The annotations provided in this bibliography were prepared only to serve as an information resource and are not intended to imply endorsement or approval for use in Chapter 1 evaluations.

Clearly there are many other tests and measurement techniques which could have been discussed in this section; keep in mind that the ones discussed here are only examples of the potential choices available for evaluation purposes.

Prior to making a final determination regarding the choice of instrumentation, the actual instruments and related publications should be carefully examined. Review copies or specimen sets of most instruments are typically available from the publishers for just this purpose. The appendices which follow this section will also be of help in making a final selection.

Appendix 5-A and 5-B, respectively, provide rating scales for selecting criterion- and norm-referenced tests. Each rating scale provides a series of questions which, when asked in relation to a specific test, will help determine whether that test is an appropriate choice for the evaluation activities. When using these rating scales it should be kept in mind that no one test will be perfect for the program evaluation activities, but some tests will be better than others.

Appendix 5-C provides a process by which a test review team can look more closely at the items in a test in order to determine how well the test measures the program's objectives. Upon completion of this process the user will be able to compare this information across tests to help to determine which test best matches the program objectives.

Appendix 5-D provides some guidelines for determining when to test out-of-level. There will be occasions when the person planning the test administration feels that the publisher's recommended test level may not be appropriate for the student(s) taking the test. If there is some question as to whether the test level will be too easy or too difficult, out-of-level testing should be considered.

Finally, Appendix 5-E provides a test administration checklist which can be used to ensure that the actual testing goes as smoothly as possible and is done correctly. Following incorrect test administration procedures can result in test data which may not be an accurate reflection of the student's scores. Therefore, following the appropriate testing procedure is very important to the evaluation process.

List of Annotations

Adult Basic Learning Examination, 1967-74

Attitude Toward School, Rev. Ed., 1972

California Achievement Tests, 1977-78

Comprehensive Assessment Program Achievement Series, 1980

Comprehensive Tests of Basic Skills, 1981-82

DIAGNOSIS: Mathematics Level B

Gates-MacGinitie Reading Tests, 1978

Instructional Objectives Exchange

Iowa Tests of Basic Skills, 1978

Mathematics In Our World, Second Edition, 1981

Measures of Self Concept, Rev. Ed., 1972

Metropolitan Achievement Tests, Instructional Battery, 1978-79

Metropolitan Achievement Tests, Survey Battery, 1978-79

Northwest Evaluation Association Item Banks

PRISM

Reading Yardsticks, 1982

Sequential Tests of Educational Progress, Series III, 1979

SRA Achievement Series, 1978

Stanford Achievement Test, 1982

Stanford Diagnostic Reading Test, 1973-78

Stanford Diagnostic Mathematics Test, 1976-78

Tests of Adult Basic Education, 1976

Wide Range Achievement Test, 1978

ADULT BASIC LEARNING EXAMINATION
(ABLE), 1967-74

Achievement: Adult
(Norm-Referenced)

ABLE is a battery of tests developed to measure the level of achievement among undereducated adults. The tests were designed to assess the knowledges and skills commonly associated with basic education or functional literacy. Although test content is adult-oriented, ABLE may be used in a variety of settings to assess achievement from the primary grades to the secondary level.

The three levels of ABLE, each available in alternate Forms A and B, measure achievement typical of grade performance from first through twelfth grade. Level I is designed for achievement levels in grades 1-4; Level II, in grades 5-8; and Level III; in grades 9-12.

Each level of ABLE consists of four tests: Vocabulary (which requires no reading), Reading, Spelling, and Arithmetic, which includes Computation and Problem Solving (dictated at Level I). Subject matter centers on aspects of practical life, such as community, family, and job. ABLE administration results can reveal the comparative strengths and weaknesses of individuals; however, they are not intended to provide indepth diagnostic information for instructional purposes.

All tests for each level and form are published in a separate booklet. The ABLE test booklets, smaller in size than conventional test booklets, are color coded and include practice items. The number of test items for each test or subtest vary by level: for Level I, the number of items ranges from 20-30; for Level II, from 20-58; and for Level III, from 42-60. The response mode for Levels I and II is varied; all items in Level III are designed in multiple-choice format. Levels I and II are available in both hand- or machine-scorable test booklet editions. Level III requires the use of a separate answer sheet which may be scored by hand or machine.

The entire battery of tests for each level can be administered in approximately two hours. However, each of the subtests may be administered in a single session. Administration time for Level I and II tests ranges from 20-25 minutes and for Level III, from 42 to 60 minutes. If more than one test is given in a single session, a rest period of 10-15 minutes should be scheduled between tests.

SelectABLE, a short 45-item screening test is available to help determine which level of ABLE is most suitable for use with an individual. The test, which covers both verbal and numerical concepts, is untimed but takes about 15 minutes to administer. The screening test and an additional ABLE test or subtest can be administered in the same testing session. Directions for test administration are provided in the ABLE Handbook that accompanies each test level. An ABLE Group Record form (one for each level) is available for recording students' scores.

In Levels I and II, the number of items correct for each test or subtest is converted to a grade score. The grade norms were established in 1966 by equating ABLE with the Stanford Achievement Test, 1964 edition. The grade norms are based on the performance of a sample of approximately 1,000 students per grade in grades 2-7, drawn from four school systems in four states. The grade norms provide a rough indication of individual performance and suggest the level of instructional materials to be used. Split-half reliability coefficients, corrected by the Spearman-Brown formula, for four research groups (students from grades 3 and 4, Job Corps enrollees, and adult basic education students in Hartford-New Haven, ranged from .73 to .98 for Level I and from .60 to .96 for Level II.

Percentiles and stanines for Level III were obtained in 1970 by equating ABLE with the Stanford Achievement Test: High School Battery, 1965-66. Reliability coefficients (KR-21), obtained from two school groups and five adult groups, ranged from .81 to .96.

ABLE is available from The Psychological Corporation, 757 Third Avenue, New York, New York 10017, (Phone: 212-888-3500) or from the publisher's regional offices. The 1982 catalog price listed for a specimen set is $5.25 for each level; and a package of 35 test booklets is $32.50 for Level I, $33.75 for Level II, and $34.50 for Level III.

ATTITUDE TOWARD SCHOOL, REV. ED., 1972 *Affective: K-12*

This 183-page collection of measurable objectives and related assessment instruments prepared by the Instructional Objectives Exchange (IOX), is devoted entirely to attitude toward school. The collection published in paperback book format, contains complete tests, along with description and rationale, directions for administration, and scoring guides.

The affective measures are intended for use in pretest/posttest evaluations of programs designed to improve student attitude toward school. From among 42 clearly defined objectives and related measures, users may select those which they consider to be appropriate for their instructional settings. Local modifications to the measures may be made if particular items are considered inappropriate. Items may be deleted, modified, or added. However, care must be taken that changes are consistent with the objective to be measured. The measures are designed to be used for assessment of group attitude only and not for individual assessment.

The attitude toward school objectives and assessment measures are arranged into three grade levels: Primary (K-3), Intermediate (4-6), and Secondary (7-12). The measures focus on five dimensions of attitude toward school: teachers, school subjects, learning, school social structure and climate, peers, and general orientation toward schooling independent of a particular school. The measures include three types: direct self report instruments which solicit student reaction in a direct question-answer format; inferential self report measures which permit inferences based on indirect stimuli questions; and observational indicators which permit inferences based on direct observation of student behavior. An overview of the measures included in the collection follows.

Direct Self Report Measures

School Sentiment Index (Primary/Intermediate/Secondary)

SSI, an omnibus inventory available in a separate version for each test level, assesses five dimensions of attitude toward school: teacher, school subjects, school social structure and climate, peer, and general. Students exhibit favorable attitudes by indicating agreement with statements that reflect positive perceptions, and disagreement with statements that reflect negative aspects of the various dimensions. The SSI includes 37 items for the Primary level, 81 items for the Intermediate level, and 82 items for the Secondary level. Administration time is approximately 10-15 minutes, 20-30 minutes, and 15-20 minutes for the Primary, Intermediate, and Secondary levels, respectively. It is recommended that the SSI be administered by someone other than the teacher to minimize a bias effect on the students' responses.

Scores available include a single global score and subscale scores for positive attitude toward school. Reliability coefficients for SSI total scores are .87 (test/retest) and .72 (KR-20) for the Primary level; .83 (test/retest) and .80 (KR-20) for the Intermediate level; and .49 (test/retest) and .88 (KR-20) for the Secondary level. Test/retest

108

coefficients for the SSI subscales across all levels ranged from .35 to 90; KR-20 coefficients for the subscales across all levels ranged from .42 to .70.

A Picture Choice (Primary, Grades K-1)

This assessment measure focuses on a student's interest in several subject areas (language, listening and speaking, science, and aesthetics (art and music). The instrument presents 28 sets of three hypothetical activities from which the child selects those he/she would like most to do. It is assumed that relative interest in the various subject areas may be inferred from the activities which the child selects. The activities are presented both orally by the test administrator and visually in the form of a picture on the student's response sheet. Scores are obtained for each subject area. Administration time is about 20 minutes. Reliability data is not available.

A Picture Choice (Primary, Grades 2-3)

This instrument requests students select one activity that he/she would most like to do from each of 30 sets of hypothetical activities in different subject areas. Each activity is presented both orally by the test administrator and visually in the form of a picture on the student's response sheet. Administration time is approximately 20 minutes. Reliability data is not available.

Inferential Self Report Measures

Subject Area Preferences (Intermediate/Secondary)

This instrument is composed of a list of subject areas commonly taught in junior and senior high school. Each subject area is accompanied by seven-degree scale on which students mark their relative preferences. This inventory provides an index of students' relative preferences among the given subject areas. Reliability coefficients for each subject area subscale across both levels range from .53 to .86 (test/retest) and .45 to .74 (KR-20).

Imagine That (Intermediate/Secondary)

This inventory presents 10 or 11 (for the Intermediate or Secondary levels, respectively) hypothetical situations regarding teacher behavior in the following areas: mode of instruction, authority and control, and interpersonal relationships with pupils. The student selects one of four alternatives for each situation. A score is obtained by totaling the number of positive alternatives selected. Reliability coefficients for the Intermediate level are .79 (test/retest) and .62 (KR-20), and .51 (test/retest) and .58 (KR-20) for the Secondary level.

The Story (Intermediate)

This instrument asks the student to select statements from a list that would fit in a realistic story about the student's school. The items describe situations expressing perceptions of the student and his peer group. The instrument is based on the assumption that a student's perception of the peer group is a component of a general attitude toward school. Reliability coefficients are .75 (test/retest) and .68 (KR-20).

Looking Back (Intermediate)

This instrument consists of 14 statements concerning remembrance of positive feelings about school that a person might make when thinking back about school years. The total score consists of the number of positive responses made by the student. Reliability coefficients are .86 (test/retest) and .67 (KR-20).

The School Play (Intermediate)

This instrument consists of 19 sentences that state positive and negative perceptions of the structure and general climate of a school. The students are asked to select those statements that could be used to write a play about their school. The score obtained is the total number of positive statements selected. Reliability coefficients are .69 (test/retest) and .74 (KR-20).

What Would Happen (Secondary)

This instrument consists of 11 fictitious situations involving two new students at school. Students are instructed to pretend that they are writing a short story and to select from among four alternatives to each situation which describe what would probably happen to the new students at their school. The reliability coefficient is .54 (both test/retest and KR-20).

Take Your Pick (Secondary)

This measure presents students with 12 hypothetical situations, each with four alternative responses. Student scores consist of the number of alternatives selected which indicate a tendency to approach rather than to avoid learning-related activities. Administration time is about 5-10 minutes. Reliability information not available.

High School on T.V. (Secondary)

This measure includes 12 hypothetical situations relating to the school social structure and climate. The student is asked to pretend that he/she is writing a television script about his/her school and to select from among three alternatives the one which depicts the most realistic details, based on their own school experience. Students are scored according to the number of positive alternatives selected. Administration time is 5-10 minutes. Reliability coefficients are .61 (test/retest) and .54 (KR-20).

Observational Indicators

Compliance With Assigned Tasks (Primary/Intermediate)

This assessment measure includes an observation record form which lists several tasks in which students might be expected to be engaged in the classroom. The measure is based on the finding that compliance with assigned tasks is a correlate of general liking of school. Data is to be collected by an outside observer. Directions for administration and scoring are provided.

School Conduct: Compliance With School Rules (Primary/Intermediate)

This observation record form utilizes available school records regarding pupils referred to school authorities. The assessment measure is based on the concept that pupils possessing favorable attitudes toward school will tend to accept the school rules and abide by them. Directions for administration and scoring are provided.

School Tardiness (Primary/Intermediate)

Records may be kept and an average daily tardiness rate computed for a specified time period. It is assumed that students who have a positive attitude toward school will tend to arrive at school on time. Directions for administration and scoring are provided.

School Attendance (Primary/Intermediate/Secondary)

Attendance records may be observed during specified time periods. It is assumed that students who possess favorable attitudes toward school will tend to incur a minimum of absenteeism. Directions for administration and scoring are provided.

Class Attendance (Secondary)

Attendance records for individual classes may be observed. It is assumed that students who hold favorable attitudes toward specific classes or subjects will tend to incur a minimum of absenteeism from those classes. Directions for administration and scoring are provided.

Class Tardiness (Secondary)

Tardiness records may be utilized as an observation indicator. It is assumed that students who hold a favorable attitude toward their classes will incur a minimum of tardiness records in arriving in those classes. Directions for administration and scoring are provided.

Grade Level Completion (Secondary, Grades 11 and 12)

The percentage of students enrolling in school each semester who complete that semester may be computed from school records. In general, those students who leave school prior to graduation tend to be those with more negative attitudes toward high school. Directions for administration and scoring are provided.

School Conduct: Compliance With School Rules (Secondary)

Records are kept by school authorities to whom students have been referred. This observation indicator is based on the concept that students possessing favorable attitudes toward school will tend to accept the school rules and abide by them. Directions for administration of the School Conduct Record Form is provided.

Unwillingness to Transfer (Secondary)

This observation indicator involves presenting the student with an option to sign up for a possible transfer to a new class section. The holding power of a class (or school) has been found to be a correlate of positive attitude toward school. Directions for administration and scoring are provided.

The objectives and assessment measures included in this IOX collection were developed by the Instructional Objectives Exchange with support from a consortium of Title III program representatives from 40 states. The present revised version was prepared following a field test with approximately 1,230 students. Items underwent extensive subject matter reviews by subject area specialists, evaluators and teachers. Statistical analyses were conducted and test/retest and KR-20 reliability coefficients were reported.

This 183-page IOX collection of measurable objectives is available from the Instructional Objectives Exchange, Box 24055, Los Angeles, California 90025. The 1981-82 catalog price is $12.95. In addition, a set of the attitude toward school measures is available in spirit master form at the elementary level and at the secondary level. The listed price for each set is $29.95.

Region V TAC
ETS-MRO

CALIFORNIA ACHIEVEMENT TESTS (CAT), 1977-78

Norm-Referenced: Grades K-12

CAT is a nationally normed, standardized achievement test battery. It was designed to provide both norm-referenced and criterion-referenced information for educational decision-making, leading to improved instruction in the basic skills areas.

CAT includes 10 levels for the following instructional grade spans: 10 (K.0-K.9); 11 (1.0-1.9); 12 (1.6-2.9); 13 (2.6-3.9); 14 (3.6-4.9); 15 (4.6-5.9); 16 (5.6-6.9); 17 (6.6-7.9); 18 (7.6-9.9); and 19 (9.6-12.9). Two alternate forms (C and D) are available for Levels 13-19; Levels 10-12 are available in Form C, only. Test include: Prereading (Listening for Information, Letter Sounds, Letter Forms, Letter Names, Visual Discrimination, and Sound Matching); Reading (Phonic Analysis, Structural Analysis, Vocabulary and Comprehension); Language (Expression and Mechanics); Spelling; and Reference Skills.

Test administration time across levels ranges from 45-96 minutes for Reading; 28-60 minutes for Mathematics; and 12-38 minutes for Language. Midpoint reference dates for test administration are November 3 and May 4. Information on test development, validity, and reliability, is reported in Technical Bulletin 1.

CAT features include: Locator Tests; Practice Tests; expanded standard score for use in functional out-of-level testing; interpolated percentile rank and NCE scores within the compliance period for Model A1; standardized directions and timing for simultaneous administrations of different test levels; Scoreze (self-scoring) Answer Sheets; Student Diagnostic Profile Sheets for recordingindividual results, including Objectives Mastery Scores; Class Record Sheets for recording group results; and the Class Management Guide, which provides follow-up instructional activities.

Test Review Kits are available from the publisher: CTB/McGraw-Hill, Del Monte Research Park, Monterey, California 93940 (Telephone: 408-649-8400 or 800-538-9547) or the publisher's regional offices. The price listed in the publisher's 1982 catalog for a Test Review Kit is $10.35 for Primary (K-3), or Intermediate (4-6), or Advanced (7-12), and $19.50 for all grades; packages of 35 reusable test booklets for each level are $18.20 for Reading or Mathematics and $32.90 for the complete battery.

COMPREHENSIVE ASSESSMENT PROGRAM
ACHIEVEMENT SERIES
(CAP ACH), 1980

Norm-Referenced: Grades PreK-12

CAP ACH is a battery of nationally normed, standardized achievement tests based on a comprehensive set of objectives. CAP ACH was designed to provide an evaluation system to yield both norm-referenced and criterion-referenced information in the basic skills areas of reading, mathematics, and language. In addition, some levels measure student performance in other areas.

The battery is comprised of 11 test levels. Suggested instructional grade ranges for each CAP ACH level are: 4 (Pre K-K.5); 5 (K.0-1.5); 6 (1.0-2.5); 7 (2.0-3.5); 8 (3.0-4.5); 9 (4.0-5.5); 10 (5.0-6.5); 11 (6.0-7.5); 12 (7.0-9.5); 13 (9.0-11.5) and 14 (11 and 12). Two parallel forms (A & B) are available for Levels 7-12. Levels 4-6 are available in one form only. Tests include: Reading (Vocabulary and Comprehension); Word Attack; Mathematics (Concepts, Computation, and Problem Solving); Language (Spelling, Capitalization and Punctuation, and Grammar); Reference and Study Skills; Writing; Science, and Social Studies.

Test administration time across levels ranges from 20-85 minutes for Reading; 20-65 minutes for Mathematics; and 20-25 minutes for Language. Empirical midpoint reference dates for test administration are October 15 and April 23. Information on test development, validity, and reliability is reported in the Technical Manual, Forms A and B

CAP ACH features include: Locator Tests; Practice Tests; expanded standard score for use in functional out-of-level testing; and the Pupil Record and Class Analyzer forms for recording test administration results.

Test Review kits are available from: Scott, Foresman Test Division, 1910 East Lake Avenue, Glenview, Illinois 60025 (Phone: 312-729-3000). The price listed in the publisher's 1982 catalog is $9.65 for combined levels 4-8 or 9-12 or 13-14; and $24-31.20 for a package of 35 reusable test booklets for each of levels 9-14.

114

COMPREHENSIVE TESTS OF BASIC SKILLS (CTBS), 1981-82

Norm-Referenced: Grades K-12

CTBS is a nationally normed, standardized achievement test battery, developed to measure achievement in the basic skills areas included in state and district curricular. In addition, some CTBS levels measure student performance in other areas. The tests were designed to provide both norm-referenced and criterion-referenced interpretation.

CTBS includes 10 levels for the following instructional grade spans: A (K.0-K.9); B (K.6-1.6); C (1.0-1.9); D (1.6-2.9); E (2.6-3.9); F (3.6-4.9); G (6.6-8.9); H (6.6-8.9); J (8.6-12.9) and K (11.0-12.9). Levels A-C are available in Form U only; Level D and above are available in alternate forms U & V. Tests include: Reading (Reading Vocabulary, Reading Comprehension, and Oral Comprehension); Visual Recognition; Sound Recognition; Language (Mechanics and Expression); Spelling; Reference Skills; Mathematics (Computation and Concepts & Applications); Science; and Social Studies.

Test administration time across levels ranges from 45-70 minutes for Reading; 15-64 minutes for Mathematics; and 15-59 minutes for Language. Midpoint reference dates for test administration are October 14 and April 29. Information on test development, validity, and reliability, is reported in Technical Bulletin 1.

CTBS features include: process/content classification of items; an expanded standard score for use in functional level testing; Locator Tests; Practice Tests; interpolated national percentile and NCE scores within the compliance period for Model A1; Scoreze (self-scoring) Answer Sheets; Student Diagnostic Profile Sheets for recording individual test results, including Objectives Mastery Scores; and the Class Management Guide, which presents information about interpretation and use of test results in instructional planning and provides suggestions for instructional activities.

Test Review Kits are available from: CTB/McGraw-Hill, Del Monte Research Park, Monterey, California (Phone: 408-649-8400 or 800-538-9547) or from the publisher's regional offices. The price listed in the publisher's catalog for a Test Review Kit is $10.35 for Primary (grades K-3) or Intermediate (grades 4-6) or Advanced (grades 7-12), or $19.50 for all grades; and $15.40 for a package of 35 reusable test booklets for Reading or Mathematics (Levels F-H, J and K, only).

DIAGNOSIS: MATHEMATIC LEVEL B *Criterion-Referenced: Grades 3-8*

DIAGNOSIS is an instructional support system comprised of a series of objective-based diagnostic tests in spirit master form. Objectives of the DIAGNOSIS system are to enable teachers to assess broad areas of achievement, pinpoint specific learner difficulties, and identify prescriptive materials and activities for remedial instruction.

Mathematics Level B, which spans the mathematics core curriculum for grades 3 through 8, consists of two learning Labs. Lab B1 covers whole number topics (concepts, addition and subtraction, multiplication and division, and word problems) and the easier topics in geometry and measurement. Lab B2 includes fractions, decimals, and related topics, and the more difficult geometry and measurement topics (graphs, statistics, and probability). The Labs are designed for use with instructional programs, and may also be adapted to other types of existing programs.

In implementing the Lab, teachers quickly assess students' skills and understanding of mathematics through the administration of the one-page Survey Tests. Students whose scores indicate learning difficulties on the Survey Tests are administered Probes, short one-page diagnostic tests that enable teachers to determine students' learning difficulties. Each Probe includes from one to four sections, each of which may be administered separately.

The section "Error Sources and Activities" in the Teacher's Guide is used to determine reasons for errors made by students. In addition, the Teacher's Guide presents instructional activities designed to help students overcome the various sources of their errors. The Prescription Guides list supplementary remedial materials for each Diagnosis learning objective. The guides correlate the objectives to major textbooks, workbooks, duplicating masters, and the publisher's supplementary materials.

Alternative Forms 1 and 2 are available for each test. For each form, there are eight Survey Tests and 10 diagnostic Probes in Lab B1 and seven Survey Tests and 10 Probes in Lab B2. Lab B1 also provides Tests of Basic Skills, in spirit master form, for use in determining whether students have adequately learned the basic facts of addition, subtraction, multiplication, and division.

The Labs include various other components. Answer Keys for the Probes and Survey Tests are printed on cards. To facilitate scoring, the answers printed on the cards are aligned with students' answers on the Probes and Survey Tests. The Student Record Sheet provides a profile of individual student progress and the Class Chart provides a profile of progress of the entire class with respect to objectives. The Class Chart may also be used for organizing small instructional groups and for managing individualized instruction.

A review set of Diagnosis: Mathematics B is available for $6.40 from the publisher: Science Research Associates, Inc., 155 Wacker Drive, Chicago, IL 60606 (Phone: 800-621-1664). The price for Lab B1 or B2 is $95 and for both Labs is $160, as listed in the publisher's 1982 catalog.

GATES-MacGINITIE READING
TESTS, 1978

Norm-Referenced: Grades 1-12

The second edition of the Gates-MacGinitie Reading Tests is comprised of a battery of nationally normed, standardized achievement tests. The tests were developed to measure achievement in reading and to provide guidelines for organizing and evaluating both individual and class instruction.

The battery includes seven levels (Basic R and A-F) and three forms. Forms 1 and 2 are available for all levels, and Form 3 is available for Levels D and E, only. Test levels are in-level for the following instructional grade ranges: Basic R (1.0-1.9); A (1.5-1.9); B (2); C (3); D (4-6); E (7-9); and F (10-12). Across levels A-F, tests include Vocabulary and Comprehension. Basic R includes four subtests (Letter Sounds, Letter Recognition, Vocabulary, and Comprehension, and a cluster of items categorized as Miscellaneous). Normative information on Basic R subtests is given descriptively—high, medium, and low.

Test administration time for Levels A-F is 20 minutes for Vocabulary and 35 minutes for Comprehension. There is no time limit for Basic R, however, total testing time is approximately 65 minutes. Empirical midpoint reference dates for test administration are: October 15 for all levels (except Level A for grade 1), February 15 for Level A for grade 1; and May 15 for all levels. Information on test development, validity, and reliability, is included in the Teacher's Manual (published separately for each level) and the Technical Summary.

Features of the Gates MacGinitie Reading Tests include: an expanded standard score (except Basic R) for functional out-of-level testing; supplementary out-of-level norms (except Level F); Decoding Skills Analysis, included in the Vocabulary tests for Levels A and B, with a form to help organize information; uniform directions and time limits that allow simultaneous administration of more than one level (A-F) within the same classroom: Teacher's Manual, which provides information on scoring, interpretation and use of scores, and also, for Level C, guidelines for error analyses of the Vocabulary test results; Self-Scorable Answer Sheet; and the Class Summary Record Sheet form.

An Examination Kit is available from the publisher: The Riverside Publishing Company, Three O'Hare Towers, 8420 Bryn Mawr Avenue, Chicago, Illinois 60631 (Phone: 800-323-9540 or 312-693-0040 in AL, HI and IL). As indicated in the publisher's catalog, the price of an Examination Kit for each level is $2.28 and $12.36 for a package of 35 hand-scorable test booklets for Levels R-E (Levels D, E, and F may be used with separate answers sheets).

Region V TAC
ETS-MRO

The IOX has three types of criterion-referenced materials available: the IOX Basic Skill System, Test/Practice Exercise Sets, and Measurable Objectives Collections.

IOX BASIC SKILL SYSTEM

The IOX Basic Skill Tests, included in this instructional system, are criterion-referenced measures of minimal competency in reading, writing, and mathematics. At the elementary level, tests are designed for minimal competency assessment at the end of grades 5 or 6, and at the secondary level, as high school graduation minimal competency measures. Each test has two alternate forms (A and B). Administration time for each test is approximately 45 minutes.

At the elementary level, the Reading test includes seven basic skills: comprehending word meaning, comprehending syntax, identifying details, identifying sequences, determining main ideas, using a dictionary, and using common reference sources. The Writing test includes seven basic skills: spelling correctly, punctuating correctly, capitalizing correctly, using pronouns correctly, selecting complete sentences, and expressing ideas in writing--optional assessment of an actual writing sample. The Mathematics test includes six basic skills: performing basic calculations with whole numbers, fractions, and decimals; solving word problems requiring a single arithmetic operation and measurement units; and interpreting tables and graphs. Elementary level tests include five items for each basic skill.

At the secondary level, the Reading test has five basic skills: understanding safety warnings, completing forms and applications, using common reference sources, determining main ideas, and using documents to take action. The Writing test has four basic skills: using words correctly, checking mechanics, selecting correct sentences, and expressing ideas in writing--optional assessment of an actual writing sample. The Mathematics test has four basic skills: performing basic calculations; and solving everyday problems requiring single arithmetic operation, formulas, multiple arithmetic operations. Secondary level tests include four or five items for each basic skill.

Other components of the IOX Basic Skills System include: test/practice exercises for secondary level only, in spirit master form; Teacher's Guides, which include detailed skill descriptions, instructional guidelines, and skill supplements; Test Manual, which provides directions for test administration and information on test content, technical test development and scoring procedures; Planning Aids, which include booklets on program planning and staff development, and audio-cassette training tapes; and Basic Skill Answer Sheets, suitable for hand or machine scoring.

TEST/PRACTICE EXERCISE SETS

These are sets of criterion-referenced measures, in spirit master form, designed for use in various ways: as practice exercises in class and as homework, as diagnostic tests, progress monitoring devices, or end of instruction measures. Most sets contain 35-50 separate test/exercises which can be matched to instructional goals. Each one- or two-page practice exercise consists of 5 or 10 items which measure a specific, well-defined objective.

Each test/practice exercise set includes: a boxed set of IOX tests, available in two alternate forms; a Test Manual, which provides descriptive test information and directions for administration and scoring; a Scoring Guide; a leaflet, "Description and Use Statement," which provides information on the development of the measures and describes possible uses of the tests for instruction and evaluation; and a packet of six Classroom Management and Program Evaluation Forms, with instructions for management or evaluation implementation.

Test/Practice Exercise Sets in basic skills areas include:

READING

Word Attack Skills, Grades K-6. There are 38 tests (44 spirit masters) in alternate forms A and B. The tests cover: vocabulary, recognition of sounds and letters, and letter and word pronunciation.

Comprehension Skills, Grades K-6. There are 40 tests (59 spirit masters) in alternate forms A and B. The tests cover: reading comprehension of main idea, conclusions, sequence, and context clues in text, as well as punctuation, syntactical structures, and affixes.

LANGUAGE ARTS

Mechanics and Usage, Grades K-6. There are 38 tests (41 spirit masters) in alternate forms A and B. The tests cover: capitalization; punctuation; abbreviation; hyphenation; bibliographic form; envelope, letter, and invitation form; plurals; possessives; pronoun referents; degree forms; subject-verb agreement; irregular past participles; misplaced modifiers; and commonly confused words.

BEST COPY AVAILABLE

Word Forms and Syntax, Grades K-6. There are 42 tests (43 spirit masters) in alternate forms A and B. The tests cover: parts of speech, verb tense and time, types of verbs, complete sentences, functions of parts of speech in sentences, types of subjects and direct objects, sentence patterns, sentence transformations, and types of clauses.

Composition, Library, and Literacy Skills, Grades K-6 There are 37 tests (57 spirit masters) in alternate forms A and B. Tests cover: sentence precision, outlining, paragraph development, paragraph transition, and types of paragraphs as well as alphabetization, dictionary use, fiction and nonfiction, the Dewey Decimal System, and card catalog use. It also includes tests for imagery, figurative language, sound patterns, figures of speech, literary elements, and literary types.

MATHEMATICS

Sets and Numbers, Grades K-6. Form A contains 35 tests (38 spirit masters) which cover sets, numbers, and rational numbers.

Operations and Properties, Grades K-6. There are 40 tests (40 spirit masters in alternate forms A and B. The tests cover: addition, subtraction, multiplication, division, and combined operations.

Numerations and Relations, Grades K-6. There are 39 tests (44 spirit masters) in alternate forms A and B. The tests cover: numeration, ratios and proportions, graphs, statistics and probability, and logic.

Measurement, Grades K-6. There are 38 tests (47 spirit masters) which cover: monetary measurement; linear measurement; liquid weight, distance, time, rate, area, and volume measurement; temperature measurement; pressure, density, and concentration measurement; and scale drawings.

Geometry, Grades K-6. There are 36 tests (45 spirit masters) in alternate forms A and B. The tests cover: points, lines, planes, simple plane figures, curves, angles, parallelism, perpendicularity, triangle similarity and congruence, circles, segments, polygons, solids, constructions, formula use, and geometric symbols.

Elements, Symbolism, and Measurement, Grades 7-9. There are 43 tests (53 spirit masters) in alternate forms A and B. Tests cover: sets, integers, rational numbers, real numbers, numeration, sentences and logic, and measurement.

Geometry, Operations, and Relations, Grades 7-9. There are 48 tests (58 spirit masters) in alternate forms A and B. Tests cover: geometry (operations and properties of planes and solids), statistics, ratios and proportions, and graphs.

Materials are available from the Instructional Objectives Exchange, Box 24055, Los Angeles, California 90025. Prices listed in the 1981-82 catalog are: $2.50 for a sample test set of the IOX Basic Skill Tests--elementary or secondary level and $37.50 to $42.50 for a package of 25 reusable test booklets in one subject area; $29.95 for one form of the Test/Practice Exercise Sets or $50.00 for both forms; and $11.95 to $15.95 for an IOX Measurable Objective Collection.

IOWA TESTS OF BASIC SKILLS (ITBS), 1978

Norm-Referenced: Grades K-9

The ITBS is comprised of a series of nationally normed, standardized achievement tests: the Primary Battery, for grades K-3, and the Multilevel Edition, for grades 3-9. The tests were designed to provide comprehensive assessment of student achievement in important areas of basic skills. Also, the Multilevel Edition includes tests in other areas. The ITBS was developed to provide both norm-referenced and criterion-referenced interpretation of test results.

The ITBS Primary Battery has four levels (5-8) which are available in one form (7) only. The Multilevel Edition has six levels (9-14) which are available in two parallel forms (7 & 8). Recommended instructional grade ranges for the ten ITBS levels are: 5 (K.1-1.5); 6 (K.8-1.9); 7 (1.7-2.6); 8 (2.7-3.5); 9 (3); 10 (4); 11 (5); 12 (6); 13 (7); and 14 (8-9). Tests include: Listening; Word Analysis; Vocabulary; and Reading or Reading Comprehension; Language or Language Skills (Spelling, Capitalization, Punctuation, and Usage); Work Study Skills (Visual Materials, and Reference Materials); and Mathematics (Concepts, Problem-Solving, and Computation).

Test administration time for Levels 9-14 is: 15 minutes for Vocabulary; 42 minutes for Reading Comprehension; 52 minutes for Total Language Skills; 65 minutes for Total Work Study Skills; and 70 minutes for Total Mathematics. Test administration time for Levels 5-8 ranges from 16-25 minutes for Listening; 14-20 minutes for Vocabulary; 20-24 minutes for Word Analysis; 34-45 minutes for Reading or Reading Comprehension; 20-47 minutes for Language/ Language Skills; 49 minutes for Work Study Skills; and 25-55 minutes for Mathematics or Mathematics Skills. Empirical midpoint reference dates for test administration are: October 28 and May 2 for the Primary Battery (Levels 5-8) and October 30 and April 28 for the Multilevel Edition (Levels 9-14). Information on test development, validity, and reliability, is reported in the Preliminary Technical Summary, Teacher's Guide (one for each of Levels 5 & 6, 7 & 8, and 9-14), and the Manual for School Administrators.

Features of the ITBS include: Practice tests; uniform directions and time limits that allow simultaneous administration of more than one level of the Multilevel Edition within the same classroom; NCE norms booklet; expanded standard score for use in functional out-of-level test administration; Teacher's Guide, which provides suggestions for improvement of students' skills in areas covered by the tests; and separate ITBS percentile norms for large cities, high and socioeconomic schools, and Catholic schools;

publisher forms including: Student Report Folders--How Are Your Basic Skills, Pupil Profile Charts, and Profile Charts for Averages; and publications: How to Use the ITBS to Improve Instruction, and Research that Built the Iowa Tests of Basic Skills.

An Examination Kit is available from the publisher: The Riverside Publishing Company, Three O'Hare Towers, 8420 Bryn Mawr Avenue, Chicago, Illinois 60631 (Phone: 800-323-9540 or 312-693-0040 in AL, HI, AND IL). As indicated in the publisher's 1982 catalog, the price of an Examination Kit is $6 for the Primary Battery or the Multilevel Edition, and $39 for a package of separate test booklets for each of Levels 9-14.

MATHEMATICS IN OUR WORLD,
SECOND EDITION, 1981

Criterion-Referenced: K-8

These objective-referenced tests, all on duplicator masters, are part of an instructional system for mathematics. The tests, packaged in individual sets by grade level are available separately from the publisher. Each set includes a series of module tests; an end-of-year test; Teachers' Guide and Answer Key; and multiple-choice answer sheets.

A detailed scope and sequence chart categorizes the content areas covered in the textbooks and tests. Content areas include: Counting and Place Value; Addition/Subtraction/Multiplication/Division of Whole Numbers; Fractional Numbers; Decimals; Estimation; Measurement; Geometry & Graphing; Percent; Problem Solving & Applications; and Special Topics.

For each grade level, there is a series of instructional modules, each with specifically defined objectives. There is a test designed for each module to determine students' mastery of the facts, skills, and basic concepts of that module. Many of the tests cover more than one objective, and the number of items that relate to a specific objective vary. Also, the number of items on each test may vary—anywhere from about 6 to 26 or more, depending on the skill area. However, most tests are printed on one page. The item format is varied—some items are direct response items but most are multiple-choice, with three or four options. On some tests, optional items are included. Some tests at the early grade levels include simple line drawings depicting children's activities.

Tests for the Mathematics in Our World series are not norm-referenced, standardized tests. There are no strict time limits or formal test administration procedures required. The publisher advises that the module tests should be used primarily as a diagnostic tool for determining which concepts or skills need further development.

Components of the instructional system include: student text & teachers' edition; Duplicator Master Tests; workbook & teachers' edition; Enrichment Workbook; and the Teachers' Resource Book, which provides reproducible masters for reteaching and enrichment, in a loose-leaf format. Across levels, the sets of duplicator masters range in price from $24.15-$25.86, as indicated in the publisher's 1982 catalog. Additional information is available from the publisher: Addison-Wesley Publishing Company, South Street, Reading, Massachusetts 01869 (Phone: 617-944-3700).

Region V TAC
ETS-MRO

MEASURES OF SELF CONCEPT, REV. ED., 1972 *Affective: K-12*

This is an IOX measurable objectives collection of affective objectives and related assessment measures available in paperback book format. The book contains complete tests, along with description and rationale, directions for administration, and scoring guides. The self concept measures are intended for use in pretest/posttest evaluations of programs designed to improve or impede increasing negativism of students' self concepts. Users may select from among the 42 clearly defined objectives and related measures, those which they consider to be appropriate for their instructional settings. Local modifications to the measures may be made if certain items are considered inappropriate for a given educational setting. Items may be deleted, modified, or added. However, care must be taken that changes are consistent with the objective to be assessed. The measures are designed to be used for assessment of group attitude only and not for individual assessment.

The objectives and related measures are arranged into three grade levels: Primary (K-3), Intermediate (4-6), and Secondary (7-12). The measures focus on four dimensions of self esteem: scholastic (derived from success or failure in scholastic endeavors), peer (associated with peer relationships), family (yielded from family interactions), and general (a comprehensive estimate). Also, measures are categorized by type: direct self report measures solicit student opinion in a straightforward question-answer format; inferential self report measures permit inferences based on student response to indirect stimuli questions; and observational indicators permit inferences on the basis of direct observation of student behavior. An overview of the measures included in the book follows.

Direct Self Report Measures

Self-Appraisal Inventory (Primary/Intermediate/Secondary)

This omnibus inventory which addresses all four dimensions of self concept: scholastic, peer, family, and general, is available in a separate version for each test level. Students demonstrate positive self concepts by indicating agreement with questions that reflect positive perceptions of the self in relation to school achievement, family, peers, and self in general; and by indicating disagreement with questions that reflect negative perceptions of self in these areas.

The inventory includes 36 items at the Primary level, 77 at the Intermediate level, and 62 at the Secondary level. Scoring is obtained by assigning weights to each response as indicated in the directions. Administration time is about 20 minutes for the Primary level, 20-30 minutes for the Intermediate level, and 15-20 minutes

How About You (Intermediate)

This inventory consists of 10 items, each with three alternatives that describe a person in relation to school and school work. The student is asked to imagine he is writing an essay and to select the descriptions which best describe him/her. The item alternatives reflect a continuum of success/failure behavior or perception of self. The inventory can be administered in 5-10 minutes. Reliability coefficients are .68 (test/retest) and .57 (KR-20).

Word Choice (Secondary)

This 19-item inventory presents the student with word pairs consisting of adjectives that describe a person's general, personal, and social attributes. Each word pair contains a semantic scale of seven degrees. The instrument is based on the assumption that a person with a positive self concept would imagine that his/her peers perceive and describe him in a favorable way. Administration time is 5-10 minutes. Reliability coefficients are .86 (test/retest) and .54 (KR-20).

For All I Know (Secondary)

This inventory consists of 10 hypothetical situations which describe backgrounds for: achievement in school, scholastic integrity, confidence in school work, scholastic initiative, and others. The alternative responses include two positive behaviors and two negative behaviors in respect to these areas. The assumption is made that a student with a positive self concept will perceive him/herself as successful and confident in scholastic endeavors. Administration time is approximately 10-15 minutes. Reliability coefficients are .31 (test/retest) and .74 (KR-20).

Observational Indicators

Word Posting (Primary/Intermediate)

This assessment technique involves creating a classroom situation in which students are provided with the option of placing their work on the bulletin board or giving it to their teacher. The measure is based on the assumption that students with a positive self concept will want to display their work. Directions for administration and scoring are provided.

Perceived Approval Situation (Primary/Intermediate/Secondary)

This assessment technique involves creating a contrived classroom situation in which a student with high self concept will identify him/herself as a member of a group that has won the approval of the teacher. Directions for administration and scoring are provided.

Inferential Self Report Measures

Television Actors (Primary)

This 18-item inventory requests students to indicate their willingness to play a wide variety of roles in an imaginary television show. The assumption made is that individuals who possess a positive self concept will be willing to project him/herself into a wider variety of roles than one who has a less strong self concept. Students reply "yes" or "no" to the questions. Administration time is approximately 10 minutes. The test/retest reliability coefficient is .74 and the KR-20 reliability coefficient is .60.

The Class Play (Primary/Intermediate)

This instrument asks the student to pretend that children are to be selected for a play and to select those roles for which his/her peers would choose him/her for. The assumption made is that an individual who has a positive self concept will perceive that others would be likely to cast him/her in roles which carry a positive image. For both Primary and Intermediate levels, approximately 10 minutes are required for administration of the 20-item inventory. Test/retest reliability coefficients are .75 for the Primary level and .80 for the Intermediate level. KR-20 reliability coefficients are .60 for the Primary level and .78 for the Intermediate level.

Parental Approval (Primary/Intermediate)

This inventory attempts to assess the extent to which a child values him/herself as unconditionally accepted by his/her mother despite trivial or major misbehavior. The inventory includes 20 items on the Primary level, which takes about 10 minutes to complete, and 10 items on the Intermediate level, which takes 5-10 minutes to complete. Reliability coefficients for the Primary level are .77 (test/retest) and .55 (KR-20). For the Intermediate level, coefficients are .91 (test/retest) and .73 (KR-20).

What Would You Do? (Intermediate/Secondary)

This inventory presents 18 fictitious situations, each followed by four alternative actions. The situations focus on the following dimensions; accomodating to others, expectations of acceptance, courage to express opinions, willingness to participate, and expectation of success. The number of positive alternatives selected consititute a student's score. There are 18 items on the Intermediate level and 19 items on the Secondary level. Both levels take about 15-20 minutes to administer. Reliability coefficients for the Intermediate level are .64 (test/retest) and .58 (KR-20) and .69 (test/retest) and .78 (KR-20) for the Secondary level.

The assessment measures included in the book were developed by the Instructional Objectives Exchange with support from a consortium of Title III program representatives from 40 states. The instruments were field tested with approximately 1,230 students in California. The items underwent extensive subject matter reviews by subject area specialists, evaluators, and teachers. Statistical analyses were conducted (test/retest and KR-20) and reliabilities coefficients were reported.

This 132-page IOX collection of measurable objectives is available from: Instructional Objectives Exchange, Box 24055, Los Angeles, California 90025. The price listed in the publisher's 1981-82 catalog is $11.95.

Region V TAC
ETS-MRO

METROPOLITAN ACHIEVEMENT TESTS, INSTRUCTIONAL BATTERY (MAT/I), 1978-79 *Criterion-referenced: Grades K-9*

MAT/I consists of a series of tests corresponding to the major instructional goals of basic skills curricula that were surveyed nationwide. The tests were designed to provide information on the educational performance of individual students in terms of specific instructional objectives. Each Instructional test assesses one basic skill area. Instructional Battery tests were coordinated in content with those of the Survey Battery. The two test batteries were nationally normed and standardized together, and have certain psychometric equivalents.

MAT/I includes six test levels for the following instructional grade spans: Primer (K.5-1.4); Primary 1 (1.5-2.4); Primary 2 (2.4-3.4); Elementary (3.5-4.9); Intermediate (5.0-6.9); and Advanced 1 (7.0-9.9). Two alternate forms (J1 & K1) are available for each level. *Reading* Instructional tests include: Visual Discrimination, Letter Recognition, Auditory Discrimination, Sight Vocabulary, Phoneme/Grapheme-Consonants, Phoneme/Grapheme Vowels, Vocabulary in Context, Word Part Clues, Rate of Comprehension, Skimming and Scanning, and Reading Comprehension. *Language* Instructional tests include: Listening Comprehension, Punctuation & Capitalization, Usage, Grammar & Syntax, Spelling, and Study Skills. *Mathematics* Instructional tests include: Numeration, Geometry & Measurement, Problem Solving, Operations-Whole Numbers, Operations-Laws & Properties, Operations-Fractions & Decimals, and Graphs & Statistics. Total Language and Total Mathematics scores are available.

Test administration time across levels ranges from 4-45 minutes for the Reading tests; 10-35 minutes for the Language tests; and 20-30 minutes for Mathematics. Empirical midpoint reference dates for test administration are October 15 and April 20, for all levels. Information on test development, validity, and reliability, is reported in the *Teacher's Manual for Administering and Interpreting*-Instructional Battery, published separately for each level, and the Metropolitan Achievement Tests *Special Reports*.

Features of MAT/I include: *Practice Tests*; Instructional Reading Level (IRL), a criterion-referenced score which enables teachers to select appropriate levels for their students from among the major basal readers; expanded standard score for functional out-of-level testing; *Class Record* form; *Teacher's Manual for Administering and Interpreting*, which provides detailed information on utilization of test scores for instructional purposes; and the Metropolitan Achievement Tests *Special Reports*.

A Sampler kit is available upon request from: The Psychological Corporation, 757 Third Avenue, New York City, New York (Phone: 212-888-3500) or the publisher's regional offices. As listed in the publisher's 1982 catalog, a complete Specimen Set for each level is $5, and a package of 35 test booklets for each level is $21.75 for Reading, $14.75-$17.50 for Language Arts, and $18.75-$20.25 for Mathematics.

METROPOLITAN ACHIEVEMENT TESTS,
SURVEY BATTERY
(MAT/S), 1978-79

Norm-Referenced: Grades K-12

MAT/S is a battery of nationally normed, standardized achievement tests designed to measure performance in the basic skills areas and, for some grades, in other areas. Survey Battery tests were coordinated in content with those of the Instructional Battery. The two test batteries were nationally normed and standardized together, and have certain psychometric equivalents. The Survey tests also yield criterion-referenced data, including estimates of Instructional Reading Level.

MAT/S includes eight test levels for the following instructional grade spans: Preprimer (K.0-K.5); Primer (K.5-1.4); Primary 1 (1.5-2.4); Primary 2 (2.5-3.4); Elementary (3.4-4.9); Intermediate (5.0-6.9); Advanced 1 (7.0-9.9); and Advanced 2 (10.0-12.9). The Primer through Advanced 2 levels are available in two alternate forms, JS & KS; the Preprimer level is available in form JS only. Tests include: Reading (Preprimer only), Reading Comprehension, Language, Mathematics, Science, and Social Studies.

Test administration time across levels ranges from 30-45 minutes for Reading Comprehension; 50 minutes for Reading (Preprimer); and 25-40 minutes for Mathematics and Language, Empirical midpoint reference dates for test administration are October 15 and April 20, for all levels. Information on test development, validity, and reliability, is reported in the Teacher's Manual for Administering and Interpreting-Survey Battery, and the series of Metropolitan Achievement Tests Special Reports.

Features of MAT/S are: Practice Tests; Instructional Reading Level (IRL), a criterion-referenced score which enables teachers to select appropriate levels for their students from among the major basal series; expanded standard score for functional out-of-level testing; Class Record form; the Teacher's Manual for Administering and Interpreting, which provides detailed information on utilization of test scores for instructional purposes; and the Metropolitan Achievement Tests Special Reports.

A Sampler kit is available upon request from: The Psychological Corporation, 757 Third Avenue, New York City, New York 10017 (Phone: 212-888-3500) or the publisher's regional offices. As listed in the publisher's 1982 catalog, a complete Specimen Set for each level is $5, and a package of 35 test booklets for each level, for Reading or Mathematics, ranges from $12.75-$14.75.

Region V TAC
ETS-MRO

NORTHWEST EVALUATION ASSOCIATION (NWEA) ITEM BANKS

Item Bank: Grades K-12

The purpose of the NWEA Item Banks is to facilitate the construction of tailor-made tests for use in instructional planning and program evaluation. Each of the three NWEA Item Banks: Reading, Language Usage, and Mathematics, includes: a collection of test item cards, a Goal Item Catalog, a User's Manual, and a test template. The test item cards are indexed in the Goal Item Catalog, by goal and Rasch level. Items are ready for reproduction following selection and arrangement.

The Reading Item Bank contains about 1,050 items related to over 150 goals, and the Mathematics Item Bank contains about 1,580 items related to over 220 goals. The number of items and related goals included in the most recently developed Language Usage Item Bank are to be made available.

The User's Manual provides detailed information on the use of the item bank and Goal Item Catalog, including item selection and retrieval, and test construction. Other components of the NWEA item bank package include K-12 Course Goals Collections (extensive taxonomies of course goals in Language Arts, Reading and Language Usage, and Mathematics). Information on arranging for technical consultation on the use of the NWEA Item Banks is included in the resource document, Guide to Consultant Services.

The item banks were developed during seven years of research, through the cooperative efforts of educational agencies in Oregon and Washington. Items which were developed to meet local district needs were pooled, indexed to curricular goals identified in the K-12 Course Goals Collections,, linked and field tested, statistically analyzed, and Rasch calibrated. The K-12 Course Goals Collections were developed by the Tri-County Goal Development Project, a consortium of about 55 school districts in Oregon, as a resource for selecting student learning outcomes for use in educational planning and evaluation.

Further information is available from Mr. Ray K. Miller, Northwest Evaluation Association, Evaluation Assessment Co-Op, ESD 121 - 1410 South 200th Street, Seattle, Washington 94148 (Phone: 2-6-242-9400). Each complete item bank (Reading, Language Usage, or Mathematics) is available for approximately $500 (1982).

Region V TAC
ETS-MRO

PRISM is a series of microcomputer software products designed for use in schools. The name PRISM is derived from its main functions: to print tests and drills, to store records, and to manage instruction and administration in the school setting. The software is designed for use with an APPLE 2, APPLE 2 Plus, or a Radio Shack Model 3 microcomputer. The software requires a microcomputer with 48k capacity, disk drive, and a compatible printer.

PRISM includes four criterion-referenced item banks from which users may generate customized "mastery tests." Each item bank is organized by broad skill areas, specific skills within these areas, relevant instructional objectives, and related test items. To use the item bank, the user selects the area, determines the specific skills within the area, and identifies the objectives to be assessed. The user also determines the number of items to be included in the test and the number of test copies to be printed. The first printed copy of the test contains the correct responses to all items.

Each item bank is available as a multi-disk program. Accompanying workbooks include longer reading passages, diagrams, graphs, and other visual displays not reproduced by the printer. Also, the Classroom Management System (available on one diskette) provides teachers with a systematic procedure for monitoring student progress. A description of the four programs follows.

PRISM MATH 1 is a criterion-referenced item bank tied to approximately 200 instructional objectives common to mathematics curricular programs in grades 3 through 8. The item bank is divided into four major areas: Numeration, Operations, Applications, and Problem Solving. Each broad skill area includes computer-generated items carefully selected for content by subject area specialists. The item bank was built around the Los Angeles Mathematics Program (LAMP), developed by the Los Angeles Unified School District and used as part of an instruction program since 1975.

The item pool consists of computer-generated items for all Operations problems and over 3,600 stored items. The computer program is stored in 15 diskettes divided into three levels: Level C-D is for use in grades 3 and 4; Level F for grades 5 and 6: and Level G-H for grades 7 and 8. At each level, the diskettes are: Program diskette, Numeration, Operations and Problem Solving, Applications, and Classroom Management Diskettes. In addition to constructing mastery tests, the item pool allows the teacher to select the particular instructional objectives in which students need additional practice. A random problem generator can produce limitless drills for all operations areas: addition, subtraction, multiplication, and division. The package also provides teachers with a systematic procedure for monitoring student progress. PRISM MATH 1 is expected to be available by late summer of 1982.

PRISM (continued)

PRISM MATH 2 is currently in preparation. The criterion-referenced item bank will focus on the basic and applied skills generally associated with minimum competency tests. Designed for students in grades 7 and above, the program may be used to develop local minimum competency tests and to provide remedial instruction on an individualized basis. Items in MATH 2 will be ungraded.

PRISM READING 1 is a criterion-referenced, multi-disk program based on the diagnostic-prescriptive approach to the teaching of reading. READING 1 is intended for use in reading centers, resource rooms, and classrooms with any type of reading program. READING 1 is also intended for use in constructing and monitoring Individualized Educational Programs.

The item bank, which includes over 2,000 items, is divided into three reading skills areas: Word Identification, Comprehension, and Study Skills that are typically taught in reading programs at grades 3-5. Each skill is measured by 8-16 items that can be used for identifying instructional needs of students, practice and drill, and for assessing mastery. Items for the item bank were drawn from the Psychological Corporation's Skills Monitoring System for Word Identification and Comprehension. To these were added a bank of items for Study Skills. The three levels C, D, and E roughly correspond to grades 3, 4, and 5, respectively. The program is expected to be available by late summer of 1982.

PRISM READING 2 is a criterion-referenced item bank focusing on the basic reading skills associated with typical minimum competency tests. It provides locally generated minimum competency tests as well as remedial drill and instruction in the skill areas where competency does not meet the minimum requirements. The 1.900 items in the READING 2 item bank are ungraded. Important comprehension skills such as reading for facts, using context clues, identifying cause and effect, and using reference materials form a substantial portion of the item bank. READING 2 is intended for students in grades 7 and above. The program is currently in preparation.

For further information about PRISM, contact: The Psychological Corporation, 757 Third Avenue, New York, New York 10017 (Phone: 212-888-3500) or the publisher's regional offices. The listed price in the publisher's 1982 catalog is: PRISM MATH 1—$250 for Levels C-D, E-F, or G-H or $675 for all three levels. PRISM MATH 2 or READING 2 prices are not yet available.

READING YARDSTICKS (RY), 1982

(Norm-Referenced)

RY is a battery of criterion-referenced tests developed to measure student performance in the skill areas needed for reading mastery at each grade level—from kindergarten through grade 8. The tests were designed to provide diagnostic information on student mastery of specific instructional objectives. Also, norm-referenced score estimates for comparable subtests on the Iowa Tests of Basic Skills (ITBS) and the Tests of Academic Proficiency (TAP) are available through the publisher's scoring service.

Reading Yardsticks has nine test levels (6-14). One form of a single level is available for each of grades K-8. Test levels are numbered to correspond roughly to chronological age. For example: Level 6 corresponds to age 6, and Level 14 corresponds to age 14. RY test levels, available in one form only, are in-level for the following grades: 6 (K), 7 (1), 8 (2), 9 (3), 10 (4), 11 (5), 12 (6), 13 (7), and 14 (8). Mastery scores, expressed in percent of items answered correctly, are available for RY Parts (Skill Areas), Subtests, and Objectives. RY Parts and Subtests include: Discrimination; Discrimination/Study Skills (Auditory Discrimination, Visual Discrimination, and Study Skills); Matching Letters and Words; Phonic Analysis (Consonant Identification, Consonant Substitution & Variants, and Vowels); Vocabulary; Comprehension; (Literal, Interpretive, Evaluative, Language, and Life Skills); Structural Analysis (Word Parts, Consonants, and Vowels); Study Skills (Reference Material, Organizational Study Skills, and Pictorial Study Skills); and Reading Rate.

Total administration time for each level ranges from 110-227 minutes. Guidelines are provided for administering the tests in two or three testing sessions, as appropriate. The equating study to equate raw scores on RY to raw scores on comparable subtests of ITBS and TAP was conducted in March and April of 1981 with a sample of about 5,000 students in grades K-8 and 10. Empirical midpoint dates for estimated ITBS norms are October 28 and May 2 for grades K-3 (Levels 6-8) and October 30 and April 28 for grades 3-9 (Levels 9-14). The midpoint reference dates for TAP norms (grades 9-12) are October 29 and April 21. Information on test development, validity, and reliability, is included in the Technical Report.

Features of Reading Yardsticks include: the Teacher's Guide, which provides information on the structure and content of the tests, planning for test administration, interpretation of scoring service reports, and use of test results; various publisher scoring services; and a student eyesight check on Levels 10-14.

An Examination Kit is available from the publisher: The Riverside Publishing Company, Three O'Hare Towers, 8420 Bryn Mawr Avenue, Chicago, Illinois 60631 (Phone: 800-323-9540 or 312-693-0040 in AL, HI, AND IL). As indicated in the publisher's 1982 catalog, the price of an Examination Kit is $2.28 for Levels 6-8 or 9-12 or 13-14; and $26.40 for a package of 35 reusable test booklets for each of Levels 9-14.

Region V TAC
ETS-MRO

SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS, SERIES III (STEP III), 1979 *Norm-Referenced: Grades PreK-12*

STEP III is a nationally normed, standardized achievement test battery. It was designed to measure student achievement in language arts, mathematics, science, and social studies. STEP III was developed for use in program evaluation and for diagnosing instructional needs. Norm-referenced domain scores, based on subscales of items within total subject area tests, are available.

STEP III consists of 10 test levels that were targeted to the following instructional grade spans: A (Pre K-K.5); B (K.5-1.5); C (1.5-2.5); D (2.5-3.5); E (3.5-4.5); F (4.5-5.5); G (5.5-6.5); H(6.5-7.5); I (7.5-10.5); and J (10.5-12.9). Subject area tests and domains for Levels E through J include: Reading (Vocabulary in Context, Literal Comprehension, and Inferential Comprehension); Vocabulary; Writing Skills (Spelling, Capitalization & Punctuation, Word Structure and Usage, and Sentence & Paragraph Organization); Listening (Listening Comprehension, and Following Directions); Study Skills (Dictionary Usage, Library Skills, and Reference Skills); Mathematics Computation (Addition of Whole Numbers, Subtraction of Whole Numbers, Multiplication & Division with Whole Numbers, Computation with Measures, Operations with Whole Numbers; Fractions/Decimals/Percents); and Mathematics Basic Concepts (Numbers & Operations, Measurement/Geometry/Graphs, and Problem Solving); Science; and Social Studies, Parallel forms X and Y are available for Levels C through J.

Test administration time for Levels E through J tests is 40 minutes except Vocabulary, which is 20 minutes, Midpoint reference dates for test administration of Levels E through J are October 5 and May 10. Information on test development, validity, and reliability, is reported in the STEP Manual and Technical Report.

STEP features include: Locator Tests; Practice Tests; expanded standard score for use in functional out-of-level testing; NCEs; Grade Level Indicators (GLIs), a grade level score based on actual test administrations in grades at, below, and above, the targeted grade span; standardized directions and timing for simultaneous test administrations of different test levels; Self-Scoring Locator Tests Answer Sheets; Local Scoring Class Record Sheets; and the publication, "The Next STEP: A Guide to Test Taking and Test Use," which includes suggestions for follow-up instructional activities.

Specimen sets are available from: Addison-Wesley Publishing Company, South Street, Reading, Massachusetts 01867 (Phone: 617-944-3700). As listed in the publisher's 1982 catalog, the price of a specimen set for Levels E-J is $4.50; and from $11.50 to $15 for a package of 35 reusable, individual

SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS, SERIES III, (continued)

subject test booklets. A package of 35 reusable Basic Assessment Test Booklets (Reading, Vocabulary, Writing Skills, Mathematics Computation, and Mathematics Basic Concepts) is $24.50 for each level. For information after June 1983, contact the Director of Cooperative Tests and Services, Educational Testing Service, Princeton, New Jersey 08541 (Phone: 609-921-9000).

Region V TAC
ETS-MRO

SRA ACHIEVEMENT SERIES
(SRA ACH), 1978

Norm-Referenced: Grades K-12

The SRA Achievement Series is a nationally normed, standardized achievement test battery. It was designed as a general survey of educational development or performance on a sample of broad content areas representative of what is taught nationally. The tests were developed to measure broad areas of knowledge, general skills, and applications, for use in evaluation of student performance and instructional programs.

SRA ACH includes eight test levels (A-H). The recommended in-level instructional grade ranges for each level are: A (K.5-1.5); B (1.5-2.5); C (2.5-3.5); D (3.5-4.5); E (4.5-6.5); F (6.5-8.5); G (8.5-10.5); and H (9-12). Parallel forms 1 and 2 are available for all levels. Tests include: Reading (Visual Discrimination, Auditory Discrimination, Letters and Sounds, Listening Comprehension, Vocabulary, and Reading Comprehension); Language Arts (Mechanics, Usage, and Spelling); and Mathematics (Concepts, Computation, and Problem Solving). The upper test levels also include Reference Materials, Science, and Social Studies.

Test administration time across levels ranges from 48-115 minutes for Reading; 45-50 minutes for Language Arts; and 23-90 minutes for Mathematics. Empirical midpoint reference dates for test administration are October 1 and April 22. Information on test development, validity, and reliability, is reported in Technical Reports 1, 2, and 3.

Features of SRA ACH include: an expanded standard score for use in functional out-of-level testing; the User's Guide which provides information on using test results for planning instruction and communicating test results; Testtalk short informational brochures on various test and measurement topics; and a series of four sound filmstrips that provide an introduction to achievement testing for teachers and parents, and provide inservice staff training on test administration, interpretation of scores, instructional planning, and district use of score reports for examining trends, problems, and solutions.

A specimen set of SRA ACH is available from the publisher: Science Research Associates, Inc., 155 North Wacker Drive, Chicago, Illinois 60606 (Phone: 312-984-7000 or, for information and assistance, 800-621-0664, except in IL, AK, AND HI. The price listed in the publisher's 1982 catalog is $2.40 for a complete Specimen Set for each level and $19.10 for all levels, and for a package of 25 reusable test booklets, $16.75 for Reading (Level D) and $24.50 for 3R (Reading, Language Arts, and Mathematics) for Levels E-H.

Norm-Referenced: Grades 1-9

(Criterion-Referenced)

STANFORD ACHIEVEMENT TEST
(STANFORD), 1982

This new edition of the Stanford Achievement Test is scheduled to be available for the 1982-83 school year. Stanford-82 is a battery of nationally normed, standardized achievement tests designed for the comprehensive assessment of the achievement status of students in the major skills areas. The tests were designed to provide both norm-referenced and diagnostic/prescriptive information for use in instructional planning and administrative decision-making.

Stanford-82 has six levels for the following instructional grade spans: Primary 1 (1.5-2.9); Primary 2 (2.5-3.9); Primary 3 (3.5-4.9); Intermediate 1 (4.5-5.9); Intermediate 2 (5.5-7.9); and Advanced (7.0-9.9). Two parallel forms (E & F) are available for each level. Tests include: Sounds and Letters, Word Study Skills, Word Reading, Reading Comprehension, Vocabulary, Listening to Words and Stories, Listening Comprehension, Spelling, Language/English, Concepts of Number, Mathematics, Mathematics Computation, Mathematics Applications, Environment, Science, and Social Science. An optional, holistically-scored, Writing test is available at the Primary 3 through the Advanced levels. A separate Using Information score, in the domain of study and inquiry, is derived from items in several subtests.

Test administration time across levels ranges from 30-70 minutes for Reading; 70-95 minutes for Mathematics; and 30 minutes for Language. The empirical fall midpoint reference date for test administration is October 7. The spring midpoint reference date will be available at a later date. Information on the national item tryout program is reported in Stanford Special Report No. 1. Additional information on test development, validity, and reliability, will be reported in the Technical Data Report.

Features of Stanford-82 include: an expanded standard score for use in functional out-of-level testing; week-of-testing; inter polated norms within the compliance period for Model A1; and NCE scores. Publications in preparation include the Guide for Classroom Planning, and the Guide for Organizational Planning.

A Sampler kit for the Stanford Series is available upon request from: The Psychological Corporation, 757 Third Avenue, New York City, New York 10017 (Phone: 212-888-3500) or the publisher's regional offices. For each test level, a complete Specimen Set is $6 and a package of 35 test booklets for Reading or Mathematics is $16, as listed in the publisher's 1982 catalog.

Region V TAC
ETS-MRO

Norm-Referenced: Grades 1-13

STANFORD DIAGNOSTIC READING TEST (SDRT), 1973-78

(Criterion-Referenced)

The SDRT is a battery of nationally normed, standardized tests designed to measure performance of the major components of the reading process. The tests are intended for use in diagnosing pupils' strengths and weaknesses in reading, and for planning instructional strategies.

The SDRT contains four test levels for the following instructional grade spans: Red (1.5-3.5); Green (2.5-5.5); Brown (4.5-9.5); and Blue (9-13). Two parallel forms (A & B) are available. Domains and Subtests include: Comprehension (Word Reading, Reading Comprehension—Sentences & Paragraphs, Literal Comprehension, Inferential Comprehension); Decoding (Auditory Discrimination, Phonetic Analysis, and Structural Analysis); Vocabulary (Auditory Vocabulary, Word Meaning, Word Parts); and Rate (Reading Rate/Fast Reading and Scanning & Skimming).

Test administration time across SDRT levels ranges from 35-50 minutes for Comprehension; 21-70 minutes for Decoding; 20-24 minutes for Vocabulary; and 2-14 minutes for Rate. Empirical midpoint reference dates for test administration are October 8 and April 28 for the Red, Green, and Brown levels (grades 1-9) and November 8 for the Blue level (grades 9-12). Information on test development, validity, and reliability, is reported in the Manual for Administering and Interpreting, published separately for each level.

Features of the SDRT include: Progress Indicators (criterion-referenced cut-off scores, for each concept/skill domain, which indicate the need for remedial instruction; Practice Test (Red and Green levels); and Instructional Placement Report (for hand scoring); and Handbook of Instructional Techniques and Materials, published separately for each level. In-level testing is recommended by the publisher, although the SDRT has an expanded standard score.

A Sampler kit is available upon request from: The Psychological Corporation, 757 Third Avenue, New York City, New York 10017 (Phone: 212-888-3500) or the publisher's regional offices. As listed in the publisher's 1982 catalog, a complete Specimen Set for each level is $4, and a package of 35 test booklets for each level is $19 (Red, Green, and Brown) and $21 (Blue).

STANFORD DIAGNOSTIC MATHEMATICS TEST (SDMT), 1976-78

Diagnostic/Prescriptive: Grades 1-13

The SDMT is a battery of nationally normed, standardized tests, designed for diagnostic/prescriptive assessment at the beginning of an instructional sequence. Test administration results are intended to be used to identify students' instructional needs and to plan instructional strategies.

The SDMT has four test levels for the following instructional grade spans: Red (1.6-4.5); Green (3.6-6.5); Brown (5.6-8.5); and Blue (7.6-13). Two alternate SDMT forms (A & B) are available. Tests include: Number System and Numeration, Computation, and Applications.

Test administration time across the SDMT levels ranges from 25-30 minutes for Number System and Numeration; 35-40 minutes for Computation; and 30 minutes for Applications. Empirical midpoint reference dates for test administration are October 8 and April 28, for all levels. Information on test development, validity, and reliability, is reported in the Manual for Administering and Interpreting, published separately for each level.

Features of the SDMT include: Progress Indicators (criterion-referenced cutoff scores) for each concept/skill domain, which indicate the need for remedial instruction; a Practice Test for the Red level; Instructional Placement Report (for hand-scoring); and the Manual for Administering and Interpreting, published separately for each level, which provides prescriptive teaching strategies. In-level testing is recommended by the publisher, although the SDMT has an expanded standard score.

A Sampler kit is available upon request from: The Psychological Corporation, 757 Third Avenue, New York City, New York 10017 (Phone: 212-888-3500) or the publisher's regional offices. As listed in the publisher's 1982 catalog, a complete Specimen Set for each level is $4, and a package of 35 test booklets for each level is $19.

TESTS OF ADULT BASIC EDUCATION (TABE), 1976

Achievement: Adult
(Norm-Referenced)

TABE is a battery of basic skills levels developed primarily for assessment of undereducated adults but intended for use in a variety of educational settings. TABE was designed to: (1) provide information about entry level skills, (2) assist in planning appropriate instructional activities, and (3) measure growth in achievement after instruction.

TABE is available in three levels: E (Easy), M (Medium), and D (Difficult) and two alternate forms (3 and 4). TABE was derived from Levels 2, 3, and 4, of the California Achievement Tests, 1970 edition. Level E corresponds to Level 2 (grades 2.0 to 4.9), Level M to Level 3 (grades 4.5 to 6.9), and Level D to Level 4 (grades 6.5 to 8.9). The language and context of CAT items were modified to reflect adult language and experience. Also, items dealing with abstract language concepts included in CAT were kept to a minimum and applications of language skills were emphasized in TABE. As part of the development of TABE, a bias review study was conducted.

Tests for all levels include: Reading (Vocabulary and Comprehension) and Mathematics (Computation, and Concepts & Problems). Language tests (Spelling, and Mechanics & Expression) are available only for Levels M and D. Subtests are further divided into sections to permit evaluation of student performance in specific skills areas.

The complete battery of tests for each level and form is included in a separate, color-coded test booklet, 8-1/2 by 11 inches. The format of test items is varied, but most are multiple choice with four or five options. For Level E, some items are presented orally by the examiner. Across TABE levels, Total Reading includes 82-85 items; Total Mathematics, 98-117 items; Mechanics & Expression, 100-109 items; and Spelling, 32 items.

TABE includes the specially developed Practice Exercise and Locator Test, printed in the same test booklet. The Practice Exercise, consisting of 14 multiple-choice items, was developed to provide students with experience in taking tests and using a separate answer sheet. The Locator Test, a short screening measure, consisting of 20 Vocabulary items and 18 Computation items, was designed as a quick screening device to determine the appropriate TABE level for testing the student. The Practice Exercise takes approximately 20 minutes to administer and the Locator Test 15 minutes. The Practice Exercise and Locator Test can be administered and scored in less than one hour, for a class of 30.

Administration time across the three TABE levels ranges from 48-59 minutes for Reading, 59-79 minutes for Mathematics, and 51-57 minutes for Language. One proctor for every 15 students is recommended. The Examiner's Manual, one for each TABE level, includes the norms tables and also, provides directions for test administration, scoring, reporting, and interpretation of test results. Answer sheets available for each level include a Complete Battery Answer Sheet and a Scoreze (self-scoring) Answer Sheet for each subject (Reading, Mathematics, or Language).

TESTS OF ADULT BASIC EDUCATION (continued)

Scores include number correct (raw scores), expanded standard (scale) scores, and grade equivalents. The norms tables provide conversions of raw scores to scale scores and to grade equivalents, and scale scores to grade equivalents. TABE norms are based on equating data collected in Spring of 1975. The equating sample included 18,183 students, representing 25 districts in 9 states, who were administered both the shortened version of CAT-70 and TABE. Grade equivalent norms for TABE were based on the scale score. Reliability coefficients (KR-20) for Total Reading and Total Mathematics ranged from .92 to .98. Other TABE components include: Directions for Administering the Practice Exercise and Locator Test; Student Profile Sheets/Analysis of Learning Difficulties for identifying instructional needs; and the Group Record Sheet for hand recording group results. Content areas in TABE are keyed to Lessons for Self-Instruction in Basic Skills (LSI), a multi-level instructional support package available from the publsher.

A Multi-Level TABE Examination Kit is available from the publisher, CTB/McGraw-Hill, Del Monte Research Park, Monterey, California 93940 (Phone: 408-649-8400 or 800-538-9547) or from the publisher's regional offices. The price listed in the publisher's 1982 catalog is $7.25. Packages of 25 test booklets are $19.50 for each level.

WIDE RANGE ACHIEVEMENT TEST (WRAT), 1978

Norm-Referenced: *PreK-12+*

Basically a clinical-type instrument designed to assess achievement in basic skills, WRAT is frequently used as a screening test to determine the approximate instructional level of students. WRAT is a wide-ranged test normed by age. WRAT is published in two test levels: Level I is intended for use with 5- to 11-year olds and Level II is intended for those 12 years and older. Both levels, printed on the same four-page test form, include three subtests: Reading (recognizing and naming letters and pronouncing words out of context); Spelling (copying marks resembling letters, name writing, and writing single words to dictation); and Arithmetic (counting, reading number symbols, solving oral problems, and performing written computations). For Reading, Level I has 100 items and Level II has 89; for Spelling, Level I has 65 items and Level II has 51; and for Arithmetic, Level I has 63 items and Level II has 57.

The three subtests may be given in any order. The Reading subtest and the oral section of the Arithmetic subtest are administered individually; the other parts of WRAT may be administered in groups. Students respond directly to the test form. The total testing time is about 20-30 minutes. Directions for the Examiner are presented in the WRAT Manual. The reading and spelling word lists are available on both plastic cards and on tape cassette. The latter can be used to train examiners or in actual test administrations.

Three main types of scores are used to report test administration results: grade ratings, percentiles, and standard scores. For each subtest, the number of items the student answered correctly is transformed into a grade rating. (For the 1978 edition, the grade ratings are printed on an insert included in the WRAT Manual.) The grade ratings are used to enter the appropriate age level norms tables in the WRAT Manual to obtain percentiles and standard scores. The standard score is comparable to an IQ or deviation score and is used for classifying students into ability categories. The publisher cautions that grade ratings should not be used to make comparisons among individuals or groups inasmuch as WRAT is normed on age and not on grade level.

The norming study included 27 age groups (from age 5 to ages 55-64) ranging in size from 400-600. The 1978 WRAT norms were adjusted on the basis of intelligence tests which were also administered to those in the norming study. The WRAT Manual presents technical information including split-half reliability coefficients for the 1965 edition which ranged from .94 to .98 for the three subtests.

Information on obtaining a specimen set is available from the publisher: Jastak Associates, Inc., 1526 Gilpin Avenue, Wilmington, Delaware 19806 (Phone: 302-652-4990). The price of a specimen set including manual and plastic cards is $19.50 and a package of 50 test forms is $8.75, as listed in the publisher's 1981-82 catalog.

Region V TAC
ETS-MRO

Additional Criterion- and Norm-Referenced Tests

DIAGNOSTIC MATHEMATICS INVENTORY (DMI), 1975-80
(Criterion- and Norm-Referenced)

The DMI is a series of criterion-referenced tests that cover 325 objectives found in both traditional and contemporary mathematics curricula currently in use. The DMI is available in seven levels (A-G) for grades 1-7+. The tests were designed to provide diagnostic information to mathematics teachers for prescribing individual and group learning activities and to provide group mastery information to administrators for needs assessment, planning, and evaluation. The DMI is keyed to 11 text book series. CTB/McGraw-Hill, Del Monte Research Park, Monterey, CA 93940.

KEYMATH DIAGNOSTIC ARITHMETIC TEST, 1971-78
(Cirterion- and Norm-Referenced)

This is an individually administered arithmetic test for use with students in prekindergarten programs through grades 6, with no upper limit for remedial use. KeyMath is comprised of 14 subtests organized into three major areas: Content, Operations, and Applications. The test has only one wide-range level with 209 test items arranged sequentially in order of difficulty. Procedures for establishing basal and ceiling levels to determine which specific items to administer to students is provided. American Guidance Service, Publishers' Building, Circle Pines, MN 55014.

NELSON READING SKILLS TEST (NRST), 1977
(Criterion- and Norm-Referenced)

The NRST is a battery of nationally normed tests in three levels (A-C) for use in grades 3-9. The tests were designed to measure student achievement in basic reading skills. Tests include Word Meaning and Reading Comprehension, which can both be administered in one class period. Optional tests include Word Parts, which permits diagnosis of specific problems, and Reading Rate. The three NRST levels are printed in a single booklet which permits group-administered individualized testing. Self-Mark Answer Sheets allow both hand-scoring and item analysis. The Riverside Publishing Company, Three O'Hare Towers, 8420 Bryn Mawr Avenue, Chicago, IL 60631.

PEABODY INDIVIDUAL ACHIEVEMENT TEST (PIAT), 1970
(Norm-Referenced)

The PIAT is a wide-range screening test of achievement in basic skills and knowledge, designed for use in kindergarten through grade 12+. There are five subtests: Mathematics, Reading Recognition, Reading Comprehension, Spelling, and General Information. Since the 84 items are sequenced in order of difficulty, students are tested within an appropriate range of difficulty, based on a basal and ceiling level. Scoring is completed during test administration. American Guidance Service, Publishers' Building, Circle Pines, MN 55014.

PRI READING SYSTEMS (PRI/RS, 1972-77
(Criterion- and Norm-Referenced)

PRI/RS is a criterion-referenced approach to assessment and instruction incorporated into articulated instructional systems in reading and related language arts. Skill areas include: Oral Language, Word Attack and Usage, Comprehension, and Applications. PRI/RS is available in five levels which span grades K-9+. Both graded and multi-graded systems are available. Materials include keyed references to widely used basal text series. CTB/McGraw-Hill, Del Monte Research Park, Monterey, CA 93940.

STS EDUCATIONAL DEVELOPMENT SERIES (EDS), 1976
(Norm-Referenced)

The EDS is a battery of nationally normed tests designed to measure achievement and ability in grades 2-12 and also, to survey and report interests and plans in relation to test results for conseling purposes. The tests are available in five levels: Lower Primary, Uppery Primary, Elementary, Advanced, and Senior. Achievement tests included in the battery are: Reading, Language Arts, and Mathematics. Scholastic Testing Service, Inc., 480 Meyer Road, Bensenville, IL 60106.

THE 3-R'S TEST, 1982
(Norm-Referenced)

This is a battery of nationally normed tests of basic skills in 11 levels (6-18) for kindergarten through grade 12. Tests include: Reading (Vocabulary, Comprehension, and Study Skills); Language (Capitalization, Spelling, Punctuation, and Grammar); and Mathematics (Computation and Problem Solving). The 3-R's is available in different editions. Administration time for the Achievement Edition is about 100 minutes; and for the Class-Period Edition, about 40 minutes. Easy-score answer sheets are available. The Riverside Publishing Company, Three O'Hare Towers, 8420 Bryn Mawr Avenue, Chicago, IL 60631.

WOODCOCK-JOHNSON PSYCHO-EDUCATIONAL BATTERY, 1977 (Norm-Referenced)

This is a wide-range battery of nationally normed tests designed to measure achievement, ability, and interest level. There are 10 achievement tests in Reading (Letter-Word Identification, Word Attack, and Passage Comprehension), Language Arts (Dictation, Proofing, Spelling, Usage, and Punctuation), and Mathematics (Calculation and Applied Problems). Basal and ceiling rules are used in most subtests to limit the range of items that must be administered. A shorter version of the test--Woodcock Language Proficiency Battery is available. Teaching Resources Corporation, 50 Pond Park Road, Hingham MA 02043-4382.

WOODCOCK READING MASTERY TESTS (WRMT), 1973

The WRMT is a 400-item, wide-range reading achievement test which is individually administered across grades K-6+. Diagnostic and instructional implications may be derived by analysis of performance on each of the five subtests: Letter Identification, Word Identification, Word Attack, Word Comprehension, and Reading Comprehension. By establishing basal and ceiling criteria, the number of specific items to administer to individual students may be determined. American Guidance Service, Publishers' Building, Circle Pines, MN 55014.

WRITING PROFICIENCY PROGRAM/INTERMEDIATE SYSTEM (WPP/IS), 1981

WPP/IS is a criterion-referenced assessment and instructional system for the management of expository writing of students in grades 6-9. Writing skills are tested with a criterion-referenced test with 70 multiple-choice items measuring 14 objectives, and three field-tested exercises which may be evaluated by holistic or primary-trait scoring. Teachers may use the results from both types of tests to evaluate students' strengths, identify areas of need, and plan instruction. CTB/McGraw-Hill, Del Monte Research Park, Monterey, CA 93440.

Additional Inventories - Affective and Other

Burke's Behavior Rating Scales
Western Psychological Services

Career Maturity Inventory
CTB McGraw-Hill

Coopersmith Self-Esteem Inventory
University of California at Davis

Intellectual Achievement Responsibility Inventory
Fels Research Institute

Kuder Occupational Interest Survey
Science Research Associates, Inc.

Locus of Control Scale for Children
Educational Testing Service

Minnesota Vocational Interest Inventory
Psychological Corporation

Minnesota School Affect Assessment, The
Center for Educational Development

Piers-Harris Self-Concept Scale
Counselor Recordings and Tests (Vanderbilt University)

Strong-Campbell Interest Inventory
Stanford University Press

Tennessee Self-Concept Scale
Counselor Recordings and Tests (Vanderbilt University)

Vineland Social Maturity Scale
American Guidance Service

Wide Range Interest and Opinion Test
JASTAK Associates

Work Values Inventory
Houghton-Mifflin Company

Additional Item Banks

California Department of Education
c/o Dr. William Padia
721 Capitol Mall
Sacramento, CA 95814

Fountain Valley Teacher Support System
c/o Richard Zweig Associates
20800 Beach Blvd.
Huntington Beach, CA 92648

Los Angeles County Test Development Center
c/o John Martois
9300 E. Imperial Highway
Downey, CA 90242

National Assessment of Education Progress
c/o Jack Schmidt
Suite 700
1860 Lincoln Street
Denver, CO 80295

Sources for Further Information

The mental measurement yearbook. Buros, O.K. (Ed.), Highland Park, NJ: The Graphon Press.

Tests in print. Buros, O.K. (Ed.), Highland Park, NJ: The Graphon Press.

A sourcebook for mental health measures. Comrey, A. L., Backer, T. E., & Glaser, E. M., Los Angeles: Human Resource Institute, 1973.

The CSE test evaluation series. Hoepfner, R., et al, Los Angeles: Center for the Study of Evaluation, UCLA Graduate School of Education.

An omnibus of measures related to school-based attitudes. Princeton, NH: Educational Testing Service, Center for Statewide Educational Assessment, 1972.

Measures of social psychological attitudes. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan, 1973.

Appendix 5-A

A CRITERION-REFERENCED TEST RATING SCALE

Directions	Test Title, Level, and Form			
1. List in the space provided the title, level, and form of each test being rated. 2. Respond to each question using the following code: Yes = √ No = - Uncertain = 0 3. If the question is not relevant, leave the space blank.				
A. TEST CONTENT/VALIDITY 1. Are the test items clearly related to the specified test objectives?				
2. Does the set of objectives and test items measured by the test match the set of objectives taught by the program?				
3. If the test covers a variety of objectives, can a subset of selected objectives be administered?				
B. RELIABILITY 1. Is each objective measured by enough test items to reasonably determine student mastery?				
2. Is the reliability for the subset of items measuring each objective high enough?				
3. Is the reliability for the entire test high enough?				
4. For each objective, is the criterion required for mastery (i.e., the number of items answered correctly) set at a reasonable level?				
5. For the entire test, is the criterion required for mastery set at a reasonable level?				
C. STUDENT APPROPRIATENESS 1. Is the response form simple enough for the students to understand?				
2. Are separate answer sheets avoided for primary grades?				
3. Is administration time of acceptable length for the students?				
4. Will students be able to understand the test instructions?				

Appendix 5-A (Cont.)

A CRITERION-REFERENCED TEST RATING SCALE

Directions 1. List in the space provided the title, level, and form of each test being rated. 2. Respond to each question using the following code: Yes = √ No = – Uncertain = 0 3. If the question is not relevant, leave the space blank.	Test Title, Level, and Form			
5. Will students be able to understand the format of the test items?				
6. Is the reading level required by the test items appropriate for the students?				
7. Is the setting required by the test the type of setting in which students in the program function best?				
8. Are the items of an appropriate interest level for the students?				
D. ADMINISTRATIVE CONSIDERATIONS 1. Are teachers or staff who will be administering the test adequately trained and/or experienced in appropriate administration of the test?				
2. Is the cost per pupil acceptable within budget constraints?				
3. Is the administration time required for the test of acceptable length in relation to the amount of time available for testing?				
4. Is it possible to use this test for other administrative testing purposes?				
5. Can several levels of the test be administered at the same time to a group of students?				
6. Can the same form and level of the test be used for pretest and posttest?				
E. SCORING CONSIDERATIONS 1. Are the desired scoring options available?				
2. If machine scoring is desired, does the publisher offer scores and score conversions needed?				

Appendix 5-A (Cont.)

A CRITERION-REFERENCED TEST RATING SCALE

Directions	Test Title, Level, and Form			
1. List in the space provided the title, level, and form of each test being rated. 2. Respond to each question using the following code: Yes = ✓ No = - Uncertain = 0 3. If the question is not relevant, leave the space blank.				
3. For machine scoring, is the publisher's "turn-around" time acceptable for instructional purposes or reports?				
4. If hand scoring is desired, is the scoring procedure clear enough to avoid errors?				
5. For hand scoring, are the necessary score conversion processes clear enough to avoid errors?				
6. Are the tables required for scoring routinely provided by the publisher?				
7. Are results summarized in terms of objectives?				
F. NORMS - NOTE: This section is optional and applies only when the test also provides norms.				
1. Does the test have empirical norms for grade level(s) of students in the program?				
2. Are the norming groups representative of the students in the program?				
3. Are the norms fairly recent?				
4. Does the test have empirical norming dates within two weeks of the pretest date?				
5. Does the test have empirical norming dates within two weeks of the posttest date?				
6. Does the test have an expanded scale score or out-of-level norms?				
7. If test norms are based on individual administration, will that be feasible given staff time available?				
8. Will the test be administered in the same type of setting upon which the norms were based?				

Appendix 5-B (Contd.)

A NORM-REFERENCED TEST RATING SCALE

Directions 1. List in the space provided the title, level, and form of each test being rated. 2. Respond to each question using the following code: Yes = ✓ No = - Uncertain = 0 3. If the question is not relevant, leave the space blank.	Test Title, Level, and Form			
A. TEST CONTENT/VALIDITY 1. Are most program objectives measured by the test items?				
2. Are most test items taught in the program?				
3. Can a subtest be administered which specifically matches the program?				
B. RELIABILITY				
1. Is test reliability high enough?				
2. Do the subtests have acceptable reliability estimates?				
C. NORMS 1. Does the test have empirical norms for grade level(s) of students in the program?				
2. Are the norming groups representative of the students?				
3. Are the norms fairly recent?				
4. Does the test have empirical norming dates within two weeks of the pretest date?				
5. Does the test have empirical norming dates within two weeks of posttest date?				
6. Does the test have an expanded scale score or out-of-level norms?				
D. STUDENT APPROPRIATENESS 1. Is the response form simple enough for the students to understand?				

A NORM-REFERENCED TEST RATING SCALE

Directions	Test Title, Level, and Form			
1. List in the space provided the title, level, and form of each test being rated. 2. Respond to each question using the following code: Yes = √ No = - Uncertain = 0 3. If the question is not relevant, leave the space blank.				
2. Are separate answer sheets avoided for primary grades?				
3. Is administration time of acceptable length for the students?				
4. Will students be able to understand the test instructions				
5. Will students be able to understand the format of the test items?				
6. Is the reading level required by test items appropriate for the students?				
7. Is the setting required by the test the type of setting in which students in the program function best?				
8. Are the items of an appropriate interest level for the students?				
E. ADMINISTRATIVE CONSIDERATIONS 1. Are teachers or staff who will be administering the test adequately trained and/or experienced in appropriate administration of the test?				
2. If test norms are based on individual administration, will that be feasible given staff time available?				
3. Will the test be administered in same type of setting upon which the norms were based?				
4. Is the cost per pupil acceptable within budget constraints?				
5. Is the administration time required for test of an acceptable length in relation to the amount of time available for testing?				

Appendix 5-B (Contd.)

A NORM-REFERENCED TEST RATING SCALE

Directions 1. List in the space provided the title, level, and form of each test being rated. 2. Respond to each question using the following code: Yes = √ No = - Uncertain = 0 3. If the question is not relevant, leave the space blank.	Test Title, Level, and Form			
6. Is it possible to use this test for other administrative testing purposes?				
7. Can several levels of the test be administered at the same time to a group of students?				
8. Will at least two-thirds of the program occur between pretesting and posttesting?				
F. SCORING CONSIDERATIONS 1. Are the desired scoring options available?				
2. If machine scoring is desired, does the publisher offer scores and score conversions needed?				
3. For machine scoring, is the publisher's "turn-around" time acceptable for instructional purposes or reports?				
4. If hand scoring is desired, is the scoring procedure clear enough to avoid errors?				
5. For hand scoring, are the necessary scoring conversion processes clear enough to avoid errors?				
6. Are the tables required for scoring routinely provided by the publisher?				

Appendix 5-C

A SUGGESTED METHOD FOR REVIEWING TESTS IN RELATION TO PROGRAM OBJECTIVES

Following these directions is a two-part rating form which can be used to analyze test content in relation to program objectives for a particular grade span. The purpose of using this form is to help determine which of several tests best matches the program objectives in reading, mathematics, or language arts. Upon completion of this rating form, the user will be able to determine, for each test, the following:

- how many test items there are for each one of the program objectives;
- the total number of program objectives measured by the test;
- the total number of items measuring the program objectives; and
- the percentage of items on the test which measure the program objectives.

Comparing this information across tests will help to determine which test best matches the program objectives.

Directions

These directions give a step-by-step procedure for using the Test Content Review Form. Blank forms, suitable for copying, are included after the directions. An example, including a completed review form, follows the blank forms.

1. On the Test Content Review Forms (One and Two) list the Title, Level, and Form of each test which is being reviewed.
2. On Form Two, Column 1, enter the total number of test items for each test being reviewed.
3. Obtain a list of all program objectives, numbering each objective consecutively. Transfer the number and a brief one- to three-word description of each objective onto

Form One. Use as many copies of the form as necessary in order to list all relevant program objectives. THE FOLLOWING STEPS SHOULD BE DONE FOR EACH TEST BEING REVIEWED.

4. Review the test for items corresponding to program objectives. Place a tally mark in the box below an objective each time a test item is found which measures that objective. Do not indicate that a test item measures more than one objective. (That is, if a test item appears to measure more than one objective, do not place a tally mark for each objective; rather, tally only the one objective which it best measures.)

5. Once all of the test items have been reviewed, count the number of program objectives measured by one or more test items and enter this number on Form Two, Column 2.

6. Total the number of items matching the program objectives (the total number of tally marks) and enter this number on Form Two, Column 3.

7. In Column 4, enter the ratio of items measuring objectives (from Column 3) to the total number of test items (from Column 1).

8. Calculate the associated percentage for this ratio (Column 3 - Column 1) and enter this percentage in Column 5. COMPLETE STEPS 4-8 FOR EACH TEST BEING REVIEWED.

Interpreting the Results

The Test Content Review Form provides a process by which a test review team can look more closely at the items in a test in order to determine how well the test measures the program's objectives.

Most parts of this form are straightforward. After completing Form One, the user will be able to determine how many test items there are for each program objective. Clearly, a test should cover most, if not all, program objectives. The number of test items per objective is also important when comparing tests. A test with three items per objective would probably be preferable to a test with only one item per objective.

Form Two is basically a summary sheet. Column 2 indicates the total number of objectives measured by the test. Again, the better test would be one that covers the most program objectives. Column 3 indicates the total number of items measuring objectives. It would be preferable to administer a test which does not include many items unrelated to the program objectives. The best indication of this is in Column 4 -- the ratio of items measuring objectives to the total number of items (also expressed in a percentage -- a higher percentage indicates that more items are of relevance to the program objectives).

In many cases there will not be a clear-cut answer as to which is the best test; it may be a matter of trade-offs. For example, one test may assess all program objectives, but with only one item per objective. Another test may cover most of the items and with more items per objective. Keep in mind when using this rating form that the purpose is to look carefully at a test in order to determine how well the items assess the program objectives. If one test is clearly not better, then a judgment will have to be made, based on the preferences of the review team.

TEST CONTENT REVIEW FORM ONE

Test Title, Level, Form	Program Objectives						
	TALLY OF ITEMS TO OBJECTIVES						

TEST CONTENT REVIEW FORM TWO

Test Title, Level, Form	1 Total Number of Test Items	2 Total Number of Objectives Measured	3 Total Number of Items Measuring Objectives	4 Ratio of: $\frac{\text{Items Measuring Objectives (3)}}{\text{Total Number of Items (1)}}$	5 Percent = Column 4 times 100

EXAMPLE

The County Institution was in the process of selecting a test of basic mathematics skills to be used for initial assessment purposes. The test would be administered after students were placed in the program in order to help determine which skill areas needed attention. Of major concern to the review team was the fact that the students, although typically in their teens, were functioning at a much lower instructional level. Because the student's attitudes toward the test content could affect their score, the team wanted to select a test with content that would be of interest to the older student, yet assess skills at the more basic level.

After conducting a preliminary review of six tests, the selection team decided to eliminate three of the tests from consideration. One test covered content totally inappropriate to the program. The other two tested for basic skills, but the content was totally inappropriate for older students. The team then decided to review three of the tests in a more extensive manner. The results of the review process are indicated on the Test Content Review Form -- EXAMPLE. On the basis of its review of the tests, the selection team came to the following conclusions about each test:

- The CDA Test measured four of the five program objectives; each objective was measured by three or more test items; and 84% of the items measured their program objectives.
- The Denby Test measured all five of the program objectives; each objective was measured by three or more test items; and 72% of the items measured their program objectives.
- The Sequential Test measured all five of the program objectives; each objective was measured by three or more test items; and 80% of the items measured their program objectives.

Based on this information, the selection team decided not to use the Denby Math Test -- the 72% figure indicated that there were too many test items that did not measure their program objectives. With the choice narrowed down to the CDA (84%) and the Sequential (80%), they decided to go with the latter. The Sequential, although it had a lower percentage, did measure all of the program objectives, while the CDA did not include items which would assess program objective 5.

EXAMPLE

TEST CONTENT REVIEW FORM ONE

Test Title, Level, Form	Program Objectives				
	1. Add Two-Digit	2. Add With Carrying	3. Subtract Two-Digit	4. Subtract With Borrowing	5. Basic Multiplication Facts
	TALLY OF ITEMS TO OBJECTIVES				
CDA Test of Basic Math Skills Level 2, Form A	///	𝍸	///	𝍸	
Denby Math Test Level 1, Form A	///	////	///	////	////
Sequential Math Skills Level 3, Form B	///	////	///	𝍸	////

EXAMPLE

TEST CONTENT REVIEW FORM TWO

Test Title, Level, Form	1 Total Number of Test Items	2 Total Number of Objectives Measured	3 Total Number of Items Measuring Objectives	4 Ratio of: Items Measuring Objectives (3) / Total Number of Items (1)	5 Percent = Column 4 times 100
CDA Test of Basic Math Skills Level 2, Form A	19	4	16	$\frac{16}{20}$	84
Denby Math Test Level 1, Form A	25	5	18	$\frac{18}{25}$	72
Sequential Math Skills Level 3, Form B	25	5	19	$\frac{20}{25}$	80

Appendix 5-D

DETERMINING WHEN TO TEST OUT-OF-LEVEL

There are some occasions when the person planning the test administration feels that the publisher's recommended test level may not be appropriate for the student(s) taking the test. If there is some question as to whether the test level will be too easy or too difficult for the student(s), out-of-level testing should be considered. There are some guidelines which will help to determine ahead of time whether out-of-level testing should be used.

First, there is a general RULE OF THUMB for existing test scores:

- If a student gets less than 1/3 (one-third) of the test items correct (floor effect), then (s)he should probably be tested at least one level lower on the same test series.
- If a student gets more than 3/4 (three-fourths) of the test items correct (ceiling effect), then (s)he should probably be tested at least one level higher on the same test series.

There are six other steps, any of which might be taken to collect more information on when to do out-of-level testing.

1. Review existing standardized test data obtained in previous years from other Chapter 1 students in the same grade as those currently under consideration. Compare the average raw scores with the above rule of thumb. If the Chapter 1 students in the same grade demonstrated a floor or ceiling effect, this year's students may do so as well.
2. Review last year's test scores for the student(s) under consideration. Again, compare the average raw score with the rule of thumb. If the students' last year's scores indicated a floor or ceiling effect, they may have the same problem this year.
3. Review the test results obtained from the test used for student selection. Applying the previously stated rule of thumb should give an

indication as to whether the test was too difficult or too easy.

4. Use the test publisher's locator tests to determine the student's functional level.

5. Review the grade level of the instructional materials selected for each student under consideration. The level of the instructional materials selected for a particular student should give a good indication of the student's functional level. Additionally, test manuals sometimes include linkages between basal reading series and test levels.

6. Obtain teachers' judgments about the student's functional level. Include past and present teachers.

Appendix 5-E

TEST ADMINISTRATION CHECKLIST*

Directions

This checklist may be used to ensure that the test administration goes as smoothly as possible and is done correctly. Place a check mark next to each step as it is completed. Following each of these steps will result in sound test administration procedures.

1. ORDER TEST MATERIALS:

____ a) Relevant information on number of booklets, levels, and forms collected.
____ b) Delivery dates of test materials confirmed.

2. SCHEDULE TEST PERIOD:

____ a) Test schedule within two weeks of norming date.
____ b) Adequate class test time scheduled.
____ c) Test date *not* around holidays or half-days.
____ d) Test date *not* during fire drills or other distractions.

3. PREPARE TEST ROOM:

____ a) Test manual consulted to determine recommended student group size for test administration.
____ b) Desks, rather than tables, in testing room.
____ c) Adequate lighting and ventilation in room.

4. PREPARE TEST ADMINISTRATORS:

____ a) Administrator's manual studied to learn all test-specific directions.
____ b) Need for keeping accurate time emphasized.
____ c) Necessity of responding to students' questions in terms of test-format, not test-content, stressed.

---

*Adapted from the Rhode Island slidetape, "A Systematic Process for Standardized Testing."

Appendix 5-E (Continued)

5. DISTRIBUTE TEST MATERIALS:

____ a) Materials counted when received:
____ b) Transfer and storage of completed tests arranged.

6. PREPARE STUDENTS:

____ a) Students informed regarding the following areas:

- purpose of test;
- areas to be tested;
- duration of test;
- date, time, and location of test; and
- materials to bring to testing session.

____ b) Practice tests, if available, administered to all students.

7. ADMINISTER TEST:

____ a) List of students to be tested made available to each test administrator.
____ b) Necessary materials, including extra pencils and scratch paper, available.
____ c) "Do Not Disturb" sign placed on door.
____ d) Students seated away from each other.
____ e) Time allowed for student questions.
____ f) Sample items administered, if any.
____ g) Student progress monitored to ensure that answer grids are being clearly marked, that students are on correct section of answer sheet or booklet, etc.
____ h) Student questions responded to only in terms of test format.
____ i) Each section of test, or entire test, administered in time prescribed in manual.
____ j) All test materials collected and counted at end of testing session.

8. SCORE TEST:

____ a) For hand scoring, test administrator's manual read for directions on how to score test and stencils used where appropriate.
____ b) For machine scoring, all answer sheets carefully packaged and accurately marked with all necessary information.

6

RECORDKEEPING
FOR PROGRAM EVALUATION
AND MANAGEMENT

6. RECORDKEEPING FOR PROGRAM EVALUATION AND MANAGEMENT

THE RELEVANCE OF COMPREHENSIVE RECORDS

Recordkeeping is a critical aspect of the overall evaluation and management of a program, especially in situations in which students may enter or leave a program at variable times during the year and in which students may come into the program with a wide range of backgrounds, capabilities and interests. Comprehensive recordkeeping techniques can contribute significantly to a variety of areas, including:

(1) **Student Management** -- The coordination of individual and group progress through both the instructional and non-instructional activities in the program

(2) **Short- and Long-Range Planning** -- The variety of planning and evaluation activities that are conducted in order to obtain information which, when used to direct plans and changes, will result in the program becoming more responsive to student needs

(3) **Evaluation and Administrative Reporting Requirements** -- All of the activities conducted in order to determine program effectiveness, as well as those necessary to meet the administrative reporting requirements

TYPES OF RECORDS AND THEIR USES

Clearly there are many types of information which should be maintained for any program as well as a variety of methods for organizing the information. This section will identify some of the types of information which should be maintained and describe some ways in which the information can be used. The user will have to select the most appropriate types of information on which to keep records on the basis of his or her program needs. Comprehensive records are the key to conducting a variety of activities, ranging from providing information for program element descriptions to providing data for program evaluation purposes.

It should be kept in mind that some of the types of records discussed here may overlap with those required by the institution. Additionally, one should be aware of the legal requirements as to the length of time certain types of information must be maintained and requirements regarding accessibility of files.

Student Management

There are a variety of types of information which should be kept on each student, as well as on the entire group. In general, three categories of information which should be maintained include: (1) individual student files; (2) individual student progress records; and (3) group progress records. Each of these types of records can contribute vital information to the development and updating of individual student plans, the management of activities during the course of the day, and the evaluation of the program.

Individual student files should primarily contain background information on the student, such as general descriptive information, previous school experiences, participation in special programs and services, test results (achievement, interest batteries, attitudinal surveys, etc.), and other general information which will contribute to the development of the most effective instructional plan for that student. On the basis of the information in the file, the staff should be able to address the following types of questions:

- What non-instructional services, such as medical or counseling, are indicated?
- On the basis of the student's strengths, interests, etc., what short- and long-term goals might be appropriate? What specific objectives?
- What further testing is indicated?
- What social situations should be encouraged? Avoided?

A form for collecting much of the information which should be maintained within the individual student file is included in Appendix 6-A.

Clearly the individual student files will vary in the amount of information actually available. While some student files may contain virtually no information, files on students who have been in the system for a longer time period may include a wealth of data. Most important, however, is that a system for collecting and maintaining this information be developed. The files should be maintained in a central location and updated on a periodic basis. If, for example, achievement tests have been administered recently but the staff is not aware of the existence of this information, then it is of no use. Finally, because much of the information is of a confidential nature, these files should be stored in a location not openly accessible.

Individual student progress records should include a description of the student's individual plan, including goals, a list of the objectives necessary to reach these goals, information on the student's learning style and preferences, the planned daily activities and schedule for the student, any behavioral systems established with the student, and a progress chart. These records should provide the information necessary to monitor the student's activities on a daily basis, to provide a periodic review of student progress, and to make revisions in the student's program as needed.

Ideally, this information would be kept in a location easily accessible to the staff involved and to the student. Updating parts of these records, such as the progress chart, could be a responsibility taken on by the student and monitored by the instructor or an aide. For the staff, information in these records can answer the following types of questions.

- Where should the student be at this point in time?
- Is the student progressing at a satisfactory rate?
- What specific goals and objectives have been met by the student?
- What areas are causing the student difficulty?
- What types of activities seem to work best for this student? Are not most effective?

For the student, information in these records can answer a variety of questions, including:

- What objective should I be working on?
- Am I making any progress?
- What learning activity should I choose?
- What are my responsibilities?

By being involved in the maintenance of progress records, the student is encouraged to take more responsibility for his or her actions. Knowing what objectives and activities are expected and how these fit into both short- and long-term goals helps the student see how learning something that may initially seem unimportant can lead to the attainment of more relevant goals. Tracking one's own progress will, in many cases, be a motivational activity in and of itself, particularly if the student is on some kind of a reward system for demonstrating mastery of certain cognitive behaviors or other skills. Furthermore, learning to take responsibility for one's actions is a skill which will benefit the student in later school, work and life situations. Examples of forms for maintaining individual student progress information are included in Appendix 6-B.

Group progress records should include a comprehensive list of the goals and objectives taught by an instructor or, in a certain course, the activities available to teach each objective, and a master progress chart by objective. These records should be maintained regardless of whether students are grouped into classes or provided with instruction on a pullout basis. These records allow the instructor to make decisions regarding activities which may benefit more than one student, to pair students for tutoring, or to decide upon other scheduling alternatives. Specifically, group progress records may help answer the following types of questions:

- Which students would benefit from being scheduled to participate in a certain activity, such as a counseling session over a defined topic, the administration of a special test, a field trip, or special tutoring on objectives which are causing difficulty?
- Which more advanced students could be used as peer tutors and which students might benefit from being tutored?
- Are there students who might form a small group to participate in a certain activity?

Examples of group progress recording forms can be found in Appendix 6-C.

Short- and Long-Range Planning

The systematic maintenance of records will facilitate the conduct of the short- and long-range planning activities that are necessary to any program. The effectiveness of a program depends upon its responsiveness to the needs of the students in the program. Long-range planning encompasses everything that is done in order to ensure that the program best meets the needs of the students who make up the target audience. This includes specific student services in both education and service areas, staff development, curriculum development, budget allocations, building and equipment needs, and follow-up services. Short-range planning encompasses the activities carried out on a more frequent and short-term basis to meet specific needs, such as student assignments to educational and non-educational services, development or revision of specific activities, assignments, resource allocations, and all other activities that are required on a daily basis.

Many of the student management records already mentioned contribute to both long- and short-range planning activities. Obviously, the individual student files and individual student progress records can contribute the basic information necessary to schedule students for instructional and non-instructional services. Changing student needs and capabilities will indicate where new materials should be added to the curriculum or what specific new activities need to be developed. Student goals and objectives will indicate what type of staff are needed, both on a long-term and short-term basis. Individual student plans will similarly indicate the types of building and equipment resources necessary to best meet needs; for example, on a short-term basis there may be a need for more programmed textbooks that seem to be very successful with the students. In the long-range picture, changing student needs may indicate that certain vocational skills would be beneficial; these new skills may require the purchase of additional equipment or a certain type of facility. Clearly, the student management records can contribute to more than just the coordination of student activities.

There are, however, additional records which should be maintained for planning purposes, including: (1) a list of goals, objectives and activities, cross-referenced by target audience; (2) an inventory of consumable and non-consumable resources; (3) a list of outside resources and types of services offered; (4) staff files; and (5) follow-up academic, vocational, and job placement availability.

A list of goals, objectives and activities cross-referenced by target audience will facilitate the assignment of specific activities to meet individual and group needs and is also critical for coordination among courses in the instructional plan, as well as for coordination with follow-up education and work placement. Ideally, there should be a master list which identifies all goals addressed within a program. Each goal should be further described by the intended target audience, the specific objectives covered, and the activities available to reach the goals and objectives. Depending upon the size of the program, the number of students and staff involved, the types of instructional and non-instructional services available, and so forth, this information may need to be classified by topic: academic, vocational, real world survival skills, staff development, etc.

In the area of short-range planning, having this information available will facilitate finding answers for the following types of situations:

- A certain student is ready to learn a specific mathematics objective. What activities are already available to teach that objective?
- A small group of students needs to learn how to behave during a job interview. Are there any simulations available to teach this content?
- Some new staff do not understand how to negotiate behavioral contracts. Are there any instructional materials on file which can help them?
- A student wants to learn how to operate a certain type of office equipment. Which objectives need to be mastered?

The availability of this type of information is possibly more important for larger programs, where informal communication is less likely. For example, a vocational instructor teaching basic mathematics required to learn a certain skill may need additional activities which cover the math skills.

While in a small program the vocational instructor may interact informally with the math instructor to obtain activities, this may not occur so easily in a larger program. The information available on goals, objectives, activities, and target audience is even more valuable for new or inexperienced staff.

In the area of long-range planning, this information will help to address the following types of questions:

- What types of academic and vocational programs are the students capable of moving into successfully?
- If a new program emphasis is going to be offered, in what areas will new activities have to be developed?
- On the basis of the most frequently used mode of instruction, what types of new equipment might be purchased in the next two years?
- How much repetition is there across skills taught in the business course and the English courses?

An inventory of consumable and non-consumable resources may seem, at first, to be a trivial type of record to be concerned with. However, without access to the necessary equipment and supplies, student and staff activities may be severely impeded, ultimately resulting in a reduction in program effectiveness. It is therefore necessary to have an updated and accurate list which indicates the status of consumable and non-consumable resources available for the conduct of the program.

For short-term planning purposes, records on resources help answer the following types of questions:

- Are the necessary answer booklets available to administer a certain test?
- Are the resources available to conduct a special class in small engine tuning?
- The equipment repair person is due tomorrow. Is there any other equipment in need of repair?
- Some funds are still available for supplies. Is there a particular item that is required right now?

For long-term planning purposes, records on resources help answer the following types of questions:

- Are there any major budget allocations that need to be made in order to upgrade or replace existing equipment?
- If a certain vocational training area is phased out during the next year, what amount of consumable resources will remain unused?
- If the program enrollment increases by five percent next year, how much in the way of additional funds would be needed for consumables?
- On the basis of available repair records, should the same brand of equipment be purchased next year or should a new vendor be selected?

A list of outside resources and types of services offered will facilitate meeting needs which cannot be met through services within the program. Depending upon the program size and internal resources, there may be areas or types of problems which cannot be adequately handled through the program. The staff must be able to identify outside resources for help in meeting these needs.

For short-term planning purposes, a list of outside resources and services offered will help solve the following types of problems:

- A new student is badly in need of dental work. What community resources may provide the needed work?
- Several teachers have indicated to the counselor that a certain student seems to be very disturbed. The counselor, who is new to the community, wants to have the student assessed by an expert. What are the options for getting the necessary testing conducted?
- An instructor wants to build a science-related field trip into an upcoming unit for a small group of students. What community resources are available for this type of activity?

For long-term planning purposes, this type of information can help answer the following types of questions:

- Does the program staff have up-to-date information regarding community resources which will provide physical or psychological assessments for students in the program?
- Are community resources being adequately used by the program staff or do they need to be made more aware of these resources?
- Are there new programs in the community which could provide the program with useful resources?

Records of staff data also provide valuable input for short- and long-term planning activities. Again, the larger the program and staff involved, the more important this task becomes. In addition to the required administrative forms, records should be maintained which describe any special interests or skills relevant to areas outside of the present assignment, courses or training sessions attended, requests for additional training, and so forth. Maintaining these types of records helps to answer the following types of questions which relate to both short- and long-term planning activities:

- If a basic computer program is begun next year, are there in-house staff capable of assuming responsibility for or participating in the program?
- There is a short-term need for an aide in the library skills course. Are there any aides with relevant background who could be freed up to participate?
- One of the students has a special interest in writing science fiction. Is there anyone on staff who could work with this student on an individualized basis?
- A training session is going to be offered on implementing behavior modification techniques for disturbed adolescents. Who from our staff would benefit most from attending?

Records describing follow-up academic, vocational and job placement opportunities are critical to the overall planning of student programs, the goals and objectives included in the program, the development of activities, the coordination among courses in the instructional program, and the coordination with outside services and programs. Staff must be aware of the follow-up programs available to students, the entry-level requirements of these programs, how these programs may be of benefit to the student, and where additional information can be located.

For short-term planning purposes, maintaining these types of records will help answer the following types of questions:

- A student has been temporarily placed in the program, but really needs some job training. What are some options?
- A student who is about to exit the program would really benefit from further counseling sessions. Is this service available?
- A counselor wants to provide a select group of students with a preview of academic follow-up opportunities. Who can provide this type of presentation to the students?

For long-term planning purposes, maintaining these types of records can help answer the following types of questions:

- For the student who will be placed in the academic program at the high school, what mathematics, English, and study skills would be most beneficial?
- For the student who hopes to get accepted in the electronics training program at the vocational school, what are the entry requirements and skills which should receive emphasis?
- A new work study program is being developed at the high school. Does our program teach skills which will facilitate entry into this program? What activities should be added?
- The support programs for the community job placement program are being reduced. How can our program compensate?

183

Evaluation and Administrative Reporting Requirements

Evaluation includes all of the activities conducted in order to determine the effectiveness of a program, as well as the various elements that make up that program. This includes areas such as determining the effectiveness of: the program in meeting individual student needs in cognitive and other skill areas; purchased or developed instructional materials or activities; non-instructional services; staff performance; staff training; the program in helping students to succeed in follow-up placements; budget expenditures; facility utilization; and any other program elements which warrant the determination of effectiveness. The systematic maintenance of records can contribute much in the way of information for evaluation purposes, thus simplifying the process of conducting the evaluation. Furthermore, the records maintained will provide most of the information necessary to meet the variety of administrative reporting requirements.

Many of the records already described will contribute to evaluation and reporting activities. The student management files, including the individual student files, the individual student progress records, and the group progress records can contain valuable information to help answer the following types of evaluation questions:

- Are students in the program demonstrating growth as indicated by standardized achievement tests? By mastery of program objectives? By locally developed criterion-referenced tests?
- What are the general characteristics of the audience which the program is serving?
- Are the non-instructional services responsive to student needs?
- What types of instructional approaches and materials are most preferred by the students? Least preferred?
- How effective are the behavioral management techniques implemented by the staff?
- Based on posttest results, which activities seem to be most effective in teaching the objectives? Least effective?
- Which staff members seem to work best with which types of students?

Much of the information maintained for short- and long-term planning activities can also contribute to evaluation and administrative reporting requirements. However, the information maintained in those records will more likely form the basis for background information, as much of it is descriptive in nature. For example, the goals, objectives and activities cross-referenced by target audience form the foundation for evaluating program effectiveness. The ongoing list of consumable and non-consumable resources can help determine the effectiveness of budget allocations. The list of outside resources is a starting place to determine how well the program makes use of community resources. The records maintained on staff help determine whether staff training has been effective or how staff effectiveness can be improved. Information on follow-up academic, vocational, and job placement programs available in the community can be a starting place to determine whether the program is realistic in regard to students' follow-up placement alternatives.

Naturally there are additional records which should be maintained to meet evaluation and administration reporting requirements. These records should include items such as: student attendance records; student selection procedures and documentation; results of past surveys administered to staff or students; follow-up data collected on students in academic, vocational, and job placements; interview documentation with potential employers; results of post-evaluation activities; state reports; and all other information which may form a basis for future evaluation or reporting activities. Maintaining records in these areas can help answer the following types of questions:

- Is there an unusually high absentee rate for certain types of students?
- What are the most typical characteristics of students selected for the program?
- Are students succeeding in their follow-up placements? What are their weaknesses? What are their strengths?
- What would future employees like to see in the way of vocational skills from graduates of the program?

Clearly there are many types of records which can contribute significantly to: student management; short- and long-range planning; and evaluation and administrative reporting requirements. As with any system, the implementation is time-consuming. However, once a system for maintaining the necessary records is begun, the information gained will be well worth the effort.

Person Completing Form

____

Date ________________

Appendix 6-A

INDIVIDUAL STUDENT FILE RECORD

Student: ____
(last name) (first) (middle)

Birth Date: _______________ Parent/Guardian: ________________________

Age: ____________________ Address: ________________________
____

Last Grade
Completed: _______________ Telephone: ________________________

Does the student have any special health problems, medication, etc.?

___ yes ___ no

If yes,
specify: ____

____

____

Date of entry into the Program: ______________________________

Estimated length of stay: ___________________________________

Has the student been in the Program before? ___ yes ___ no

If yes, dates: ____

Location of last placement: __________________________________

Type of program (e.g., academic, vocational, residential treatment):

____

Dates attended: ____

Who should be contacted for further information regarding this last placement (e.g., counselors, probation officer, program administrator, teachers):

Name ______________________ Name ________________________

Telephone ____________________ Telephone ____________________

Has the student received special services in the past (e.g., health care, counseling)?

____ yes ____ no ____ unsure

If yes, specify date(s) and type(s) of program: ______________________________

____

Has the student been placed in special types of programs (e.g., learning disabled, work-study)?

____ yes ____ no ____ unsure

If yes, specify date(s) and type(s) of program: ______________________________

____

____

- -

TEST/OTHER INFORMATION AVAILABLE (MOST RECENT):

Achievement Test(s):

Test Title, Form and Level: ______________________________

Date of Administration: ______________ Administered by: ______________

Area ________ Raw Score ______ Percentile ______ Standard Score ______

Area ________ Raw Score ______ Percentile ______ Standard Score ______

Area ________ Raw Score ______ Percentile ______ Standard Score ______

Area ________ Raw Score ______ Percentile ______ Standard Score ______

Test Title, Form and Level: ______________________________

Date of Administration: ______________ Administered by: ______________

Area ________ Raw Score ______ Percentile ______ Standard Score ______

Area ________ Raw Score ______ Percentile ______ Standard Score ______

Area ________ Raw Score ______ Percentile ______ Standard Score ______

Area ________ Raw Score ______ Percentile ______ Standard Score ______

Test Title, Form and Level: ______________________________

Date of Administration: __________ Administered by: __________

Area ______ Raw Score ______ Percentile ______ Standard Score ______

Area ______ Raw Score ______ Percentile ______ Standard Score ______

Area ______ Raw Score ______ Percentile ______ Standard Score ______

Area ______ Raw Score ______ Percentile ______ Standard Score ______

Psychological Assessment(s):

Name of Instrument: ______________________________

Date of Administration: __________ Administered by: __________

Scores/Conclusions: ______________________________

Name of Instrument: ______________________________

Date of Administration: __________ Administered by: __________

Scores/Conclusions: ______________________________

Interest/Attitude Survey(s):

Name of Instrument: ______________________________

Date of Administration: __________ Administered by: __________

Scores/Conclusions: ______________________________

Name of Instrument: ______________________________

Date of Administration: __________ Administered by: __________

Scores/Conclusions: ______________________________

Other Assessments (Vocational, Aptitude):

Name of Instrument: ____________________

Date of Administration: __________ Administered by: __________

Scores/Conclusions: ____________________

____

____

Name of Instrument: ____________________

Date of Administration: __________ Administered by: __________

Scores/Conclusions: ____________________

____

____

- -

Student's special interests/hobbies: ____________________

____

____

Student's job interests/goals: ____________________

____

____

How well does the student function with peers? ____________________

____

With those younger? ____________________

____

With those older? ____________________

____

With adults/authority? ____________________

____

How well does the student function independently? ______________________

____

____

In one-to-one situations? ____

____

In small groups? ____

____

In large groups? ____

____

In highly structured situations? _________________________________

____

In unstructured situations? ______________________________________

____

What are some potential methods of positive reinforcement which might be used with this student? ______________________________________

____

____

____

____

Are there certain situations/settings/types of interaction which should be avoided for this student? ______________________________________

____

____

____

What specific academic/vocational strengths does the student have? ______

____

____

What academic areas need special attention? ____________________

____

____

____

What non-academic areas (interactive skills, grooming, physical coordination, etc.), need special attention? ____________________

____

____

____

Has an Individual Student Plan been developed?

_____ Yes (attach to file)

_____ No (If not, when will one be developed?) ____________________

Individual(s) responsible for plan: ____________________

____

The following items should be included in this file:

- The student's schedule, both daily and planned special events (e.g., weekly counseling sessions).
- A list of goals/objectives.
- The Individual Student Plan.
- Student Management Plans, if developed.
- An identification of special curriculum materials/programs being used.
- Any interest surveys, student preference questionnaires, etc., administered by the program personnel.
- Examples of past work.
- Other past records/grades which will be of use in planning the student's program.

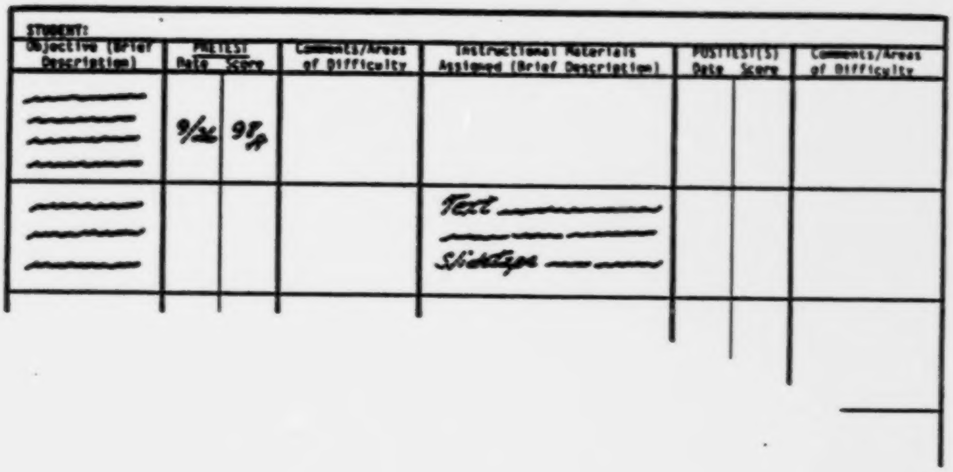
Appendix 6-B

INDIVIDUAL STUDENT PROGRESS RECORDS

As discussed in Section 6, there is a variety of information which can be maintained on individual student progress. Three forms which may be of use are included here.

Form One includes a place for the student's name and a column to briefly describe each objective. For each objective there is space to record the pretest date and score, notes regarding areas of difficulty, and a short description of the instructional materials assigned. There is also a column for recording posttest results and areas of difficulty. As an example, in the abbreviated form below it can be quickly determined that the student passed the pretest for the first objective and is now working on another objective; the assigned materials are described in the designated column. No posttest has been attempted.

INDIVIDUAL PROGRESS INFORMATION - FORM ONE

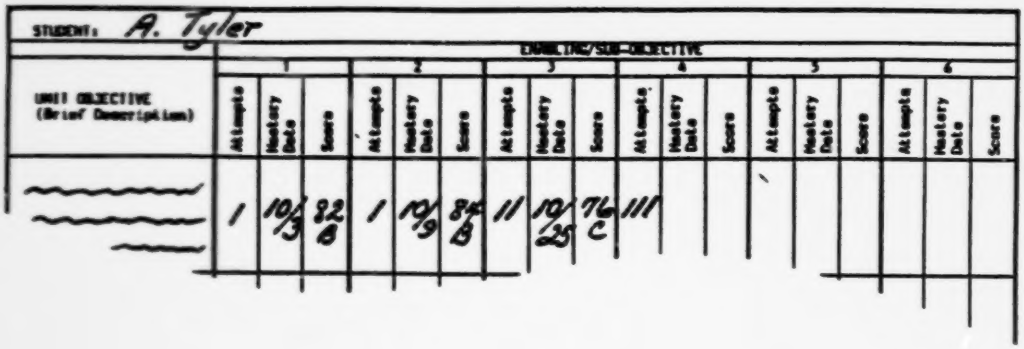


STUDENT:

Objective (Brief Description)	PRETEST Date	PRETEST Score	Comments/Areas of Difficulty	Instructional Materials Assigned (Brief Description)	POSTTEST(S) Date	POSTTEST(S) Score	Comments/Areas of Difficulty
	9/26	98%					
				Text Slidetape			

Form Two provides a method for collecting individual progress information on unit and enabling objectives. For each unit objective, a student's progress can be tracked in regard to each enabling objective, including test attempts, the mastery date, and the test score. For example, in the abbreviated form that follows, Unit Objective 1 has five enabling objectives. The student easily mastered Enabling Objectives 1 and 2, required two attempts for Enabling Objective 3, and is having difficulty with Enabling Objective 4 -- the test has been attempted three times and still has not been passed.

INDIVIDUAL PROGRESS INFORMATION - FORM TWO

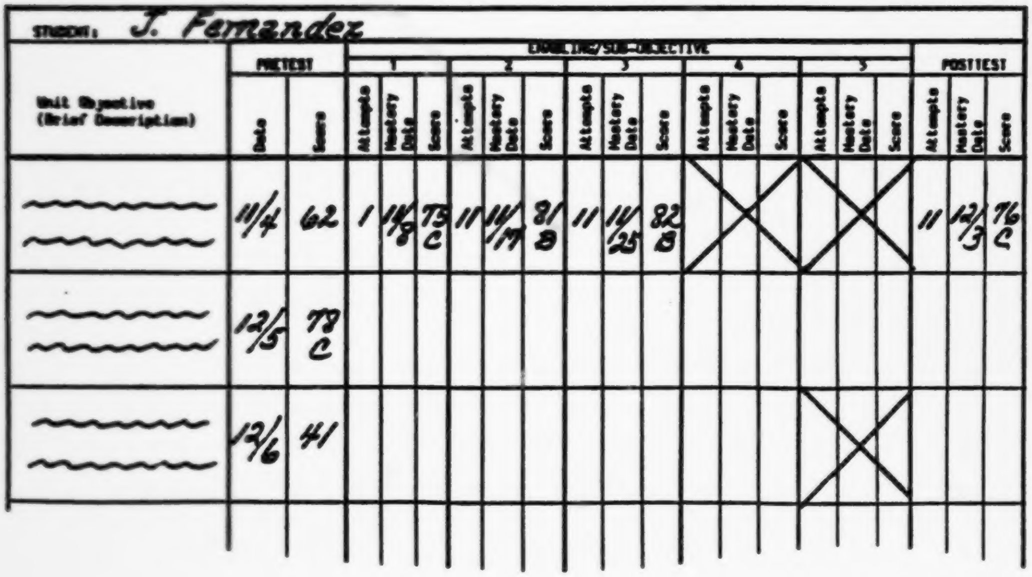


STUDENT: A. Tyler

UNIT OBJECTIVE (Brief Description)	ENABLING/SUB-OBJECTIVE 1			2			3			4			5			6		
	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score
	1	10/3	82 B	1	10/9	84 B	11	10/25	76 C	111								

Form Three provides a way to collect individual progress information on both unit objectives and enabling objectives. For each unit objective there is space to record information on both pretest and posttest scores. For each enabling objective there is a column to record test attempts, the date of mastery, and the test score. In the abbreviated form below the student did not pass the pretest for the first objective, but did go on to master the enabling objectives and the posttest. The pretest for the second unit objective was passed on the first attempt. The student did not pass the pretest for the third unit objective and is now working on the first enabling objective.

INDIVIDUAL PROGRESS INFORMATION - FORM THREE



STUDENT: J. Fernandez

Unit Objective (Brief Description)	PRETEST		ENABLING/SUB-OBJECTIVE 1			2			3			4			5			POSTTEST		
	Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score
	11/4	62	1	11/8	75 C	11	11/19	81 B	11	11/25	82 B							11	12/3	76 C
	12/5	78 C																		
	12/6	41																		

Appendix 6-B

INDIVIDUAL PROGRESS INFORMATION - FORM ONE

STUDENT:

Objective (Brief Description)	PRETEST		Comments/Areas of Difficulty	Instructional Materials Assigned (Brief Description)	POSTTEST(S)		Comments/Areas of Difficulty
	Date	Score			Date	Score	

Appendix 6-B (Cont.)

INDIVIDUAL PROGRESS INFORMATION - FORM TWO

STUDENT:	ENABLING/SUB-OBJECTIVE																	
Unit Objective (Brief Description)	1			2			3			4			5			6		
	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score

Appendix 6-B (Cont.)

INDIVIDUAL PROGRESS INFORMATION - FORM THREE

STUDENT:

Unit Objective (Brief Description)	PRETEST		ENABLING/SUB-OBJECTIVE															POSTTEST		
			1			2			3			4			5					
	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score	Attempts	Mastery Date	Score

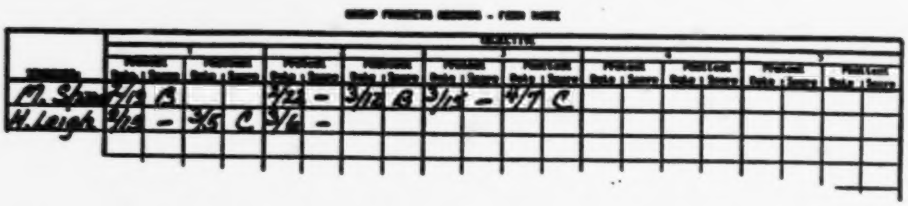
Appendix 6-C

GROUP PROGRESS RECORDS

As covered in Section 6, there is also a variety of information wihch can be maintained on groups of students. Four forms which may be of use in this area are included here.

Forms One and Two do not relate directly to tracking progress, but do provide a means for collecting information necessary to manage the activities of students, both on an individual basis and as a group. Form One is simply a method to identify, for each goal, the related objectives and instructional materials available for that objective. By using this form, a master list of all program goals, objectives and instructional materials can be maintained in one location. Form Two can be used as a master list to identify objectives cross-referenced to activities and materials available to provide instruction for each objective.

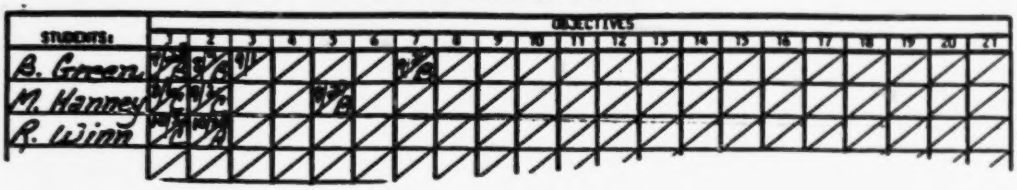
Forms Three and Four are both examples of forms which can be used to maintain records over the progress of a group of students. In Form Three, each student is listed down the left-hand column. For each objective there is a column to record the date of the student's pretest and score, as well as the posttest and score. In the abbreviated example below it can be seen that M. Shaw has passed the pretest for Objective 1 and passed the posttest for Objectives 2 and 3. The second student, H. Leigh, has passed the posttest for Objective 1, but did not pass the pretest for Objective 2.



M. Shaw	2/13	B			2/23	–	3/12	B	3/15	–	4/7	C								
H. Leigh	3/15	–	3/5	C	3/6	–														

Form Four provides a way to track progress simply by objective. Each student is listed down the left-hand column of the form. For each objective there is a space to indicate the date the test was passed and the score. In the abbreviated form below it can be seen that B. Green has completed Objectives 1, 2, 3, and 7. M. Hanney has completed Objectives 1, 2, and 5, while R. Winn has completed Objectives 1 and 2.

GROUP PROGRESS RECORDS - FORM FOUR



STUDENTS:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
B. Green	[illegible]	[illegible]	[illegible]				[illegible]														
M. Hanney	[illegible]	[illegible]			[illegible]																
R. Winn	[illegible]	[illegible]																			

Appendix 6-C

GROUP PROGRESS RECORDS - FORM ONE

Goal 1:

Objective 1.1:

Instructional Materials Available:

Objective 1.2:

Instructional Materials Available:

Objective 1.3:

Instructional Materials Available:

Appendix 6-C

GROUP PROGRESS RECORDS - FORM TWO

OBJECTIVES:	Activities/Materials																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q

Appendix 6-C

GROUP PROGRESS RECORDS - FORM THREE

STUDENTS:	OBJECTIVE																			
	1				2				3				4				5			
	Pretest		Posttest		Pretest		Posttest		Pretest		Posttest		Pretest		Posttest		Pretest		Posttest	
	Date	Score	Date	Score	Date	Score	Date	Score	Date	Score	Date	Score	Date	Score	Date	Score	Date	Score	Date	Score

Appendix 6-C

GROUP PROGRESS RECORDS - FORM FOUR

	OBJECTIVES																				
STUDENTS:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21

7

REQUIREMENTS AND RECOMMENDATIONS FOR EVALUATING CHAPTER 1 N OR D PROGRAMS

7. REQUIREMENTS AND RECOMMENDATIONS FOR EVALUATING CHAPTER 1 N OR D PROGRAMS

INTRODUCTION

This section of the handbook has been included primarily as a place to insert specific regulations and other information which is frequently referenced by the user. At a minimum, it is suggested that a copy of the Chapter 1 Law and Federal Regulations related to N or D programs be inserted in this section. Specific state policies, reporting forms, and other state-related documents should also be inserted here.

This section also presents a description of recommended information which might be collected by each program for overall summary purposes. This description is followed by two examples of reporting forms -- a long, comprehensive version (Appendix 7-A) and a short version containing information considered to be most essential (Appendix 7-B).

RECOMMENDED INFORMATION TO BE COLLECTED BY PROJECTS

Although there are no requirements for reporting evaluation results of N or D programs to the Federal government under Chapter 1 of The Education Consolidation and Improvement Act of 1981 (PL 97-35), there are some types of information which might be collected by each program. This information would be of use both to the program and to those who are interested in finding out more about that program. For example, the project that collects the suggested information will find it easier to accurately and completely describe the program to others. Similarly, should N or D project directors have an opportunity to share information about their programs, the information summarized in the following topics will provide a common base for initiating discussions.

1. Name of Institution/Facility

2. Address of Institution/Facility

3. Name of Person Completing Form

4. Above Person's Telephone Number

5. Duration of Project

A Chapter 1 project would generally be expected to be in operation for one year or less; the beginning and ending dates should be provided.

6. Number of persons under 21 years who resided at the institution/facility during the project, by age in years and last grade completed

Age and last grade completed should be determined as of the beginning of the Chapter 1 project or as of assignment to the program, if later than the beginning of the project. The number should be an unduplicated count (i.e., each person is counted only once no matter how many times that person might be assigned to that institution/facility during the Chapter 1 project.

7. Number of persons under 21 years who participated in the project, by age in years and last grade completed

Age and last grade completed should be determined as of the beginning of the Chapter 1 project or as of assignment to the project, if later than the beginning of the project. The number should be an unduplicated count (see item 6) and, therefore, should be no larger than the number of residents counted in Item 6.

8. Number of project participants by racial/ethnic group

The suggested racial/ethnic groups are defined below. The total should equal the grand total in Item 7.

- American Indian or Alaskan Native - a person having origins in any of the original peoples of North America, and who maintains cultural identification through tribal affiliation or community recognition.

- Asian or Pacific Islander - a person having origins in any of the original peoples of the Far East, Southeast Asia, the Indian subcontinent, or the Pacific Islands. This area includes, for example, China, India, Japan, Korea, the Phillippine Islands and Samoa.

- Black, not Hispanic - a person having origins in any of the black racial groups of Africa.

- Hispanic - a person of Mexican, Puerto Rican, Cuban, Central or South American or other Spanish culture or origin, regardless of race.

- White, not Hispanic - a person having origins in any of the original peoples of Europe, North Africa, or the Middle East.

9. Number of project staff by job classification

For each job classification, record the number of persons paid at least in part, by the Chapter 1 project in terms of full-time equivalent (FTEs). To calculate the number of FTE staff members in a job classification, determine the number of hours each person in that job classification worked per week. Add these numbers together and divide by the number of hours per week that represents "full-time" for that job classification. If an individual works in a Chapter 1 project for more hours than are paid by Chapter 1 funds, count the FTE in terms of hours worked rather than hours paid by Chapter 1. The job classifications are defined as follows:

- Administrative Staff - a person whose primary assignment is to direct staff members or manage the Chapter 1 project and its supporting services within an operating unit or facility (e.g., project directors, coordinators).

- Teachers - staff members who instruct Chapter 1 students.

- Teacher Aides - staff members who assist a teacher with routine activities associated with teaching and those activities requiring minor decisions regarding students, such as monitoring, coordinating exercises, operating equipment, and maintaining records.

- Other - includes curriculum specialists, support service staff (e.g., social work, guidance, psychological counseling, health, nutrition), clerical staff and others not included above. Specify each "other" job classification and provide a separate count for each.

10. Number of project and non-project staff by job classification who received training funded by the Chapter 1 project

Non-project staff are those not paid at all by Chapter 1 funds. Job classifications are defined in Item 9. Numbers should be unduplicated counts (i.e., one person receiving training on more than one occasion is counted only once). Specify each "other" job classification and provide a separate count for each.

11. Number of project participants who received services in each project component area

For each Chapter 1 project component area in which services were provided, specify the number of participants who received that service. The number should be a duplicated count (i.e., a student should be counted once for each area in which services were received). Other instructional areas include English to limited English speaking students, vocational education, special education for handicapped, etc. Supportive Services include social work, guidance, psychological counseling, health, nutrition, student transportation, etc. Specify each "other instructional area" and each "supportive service" and provide separate counts for each.

12. Number of project participants who received component services by duration of service in months and by last grade completed

Last grade completed should be as of the beginning of the Chapter 1 project or as of assignment to the program, if later than the beginning of the project. Duration of service should be the total number of months each student received component services over the entire project. If a student's duration of service was an exact number of months, count that student in the first category containing that number of months. For example, if duration of service was exactly eight months, count that student in the 7-8 category. The grand total should equal the component count in Item 11.

13. State the question(s) addressed for the purpose of evaluating the short-term effectiveness of the component.

    Refer to Sections 1 and 3 of this handbook for a discussion of purposes of evaluation and evaluation questions, respectively. Provide only those questions which concern the short-term effectiveness of the component's services here. Other questions would be more appropriately discussed and addressed in a more comprehensive, final evaluation report.

14. Check the indicator(s) of need for service used to select the participants in the component.

    Check one or more types of information used to indicate a student's need for the Chapter 1 project component's services. Test scores could include achievement tests, diagnostic tests, affective measures, etc. Teacher judgments could include skill-deficiency checklists, general referrals, estimated grade level, etc., and could be provided by non-Chapter 1 or Chapter 1 teachers. Other judgments could include referrals made by other instructional or support service staff, the students, etc. Other performance indicators could include grades, level of instructional materials, previous participation in Chapter 1 projects, etc.

15. Check the type(s) of instrument(s) used to measure the intended impact of the component.

    Check each type of measurement instrument used to assess the intended impact of the component services on the students. For example, a reading instructional component would use some measure of reading achievement level; an affective component might use a self-concept scale; while a counseling component might employ interviews or existing records.

16. On a separate sheet, provide the results of the analysis of the effectiveness of the project component using each of the instruments checked above. Include at least the following:

- a measure of the component participants' average achievement level, performance, attitude, etc., after receiving component services (e.g., an average posttest score);
- a measure of the average achievement level, performance, attitude, etc., expected of the participant had they not received component services (e.g., an average pretest score);
- the number of scores the above averages are based upon;
- a brief description/identification of each instrument;
- the type of score used in the analysis (e.g., NCE, standard score, raw score, number of objectives, etc.); and
- the date(s) when each instrument was administered or completed.

In the following appendices are the examples of the forms mentioned in the introduction to this section. The first example (Appendix 7-A) is of a long, comprehensive form for collecting program information. The second example (Appendix 7-B) is a short version for collecting information considered to be most essential.

Appendix 7-A

SAMPLE FORM FOR COLLECTING DATA FOR CHAPTER 1 N OR D PROJECTS

(Long Version)

1. Name of Institution/Facility ______________________________

2. Address of Institution/Facility ______________________________

____

3. Name of Person Completing Form ______________________________

4. Above Person's Telephone Number ______________________________

5. Duration of Project: From ______________ To ______________

6. Number of persons under 21 years who resided at the institution/facility during the project by age in years and last grade completed at the beginning of the project or at commitment if later than the beginning of the project (unduplicated count)

Grade / Age	3 or under	4	5	6	7	8	9	10	11	12	Total
10 or under											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
Total											

7. Number of persons under 21 years who participated in the project by age in years and last grade completed at the beginning of the project or at commitment if later than the beginning of the project (unduplicated count)

Grade / Age	3 or under	4	5	6	7	8	9	10	11	Total
10 or under										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
Total										

8. Number of project participants by racial/ethnic group

American Indian or Alaskan Native ____________
Asian or Pacific Islander ____________
Black, not Hispanic ____________
Hispanic ____________
White, not Hispanic ____________
Total ____________

9. Number of project staff (paid at least in part by Chapter 1 funds) by job classification (in FTEs)

Administration Staff ________.___
Teachers ________.___
Teacher Aides ________.___
Other (Specify) ____________________ ________.___
____________________ ________.___
____________________ ________.___

10. Number of project and non-project staff by job classification who received training funded by the Chapter 1 project

		Project	Non-Project
Administrative Staff		______	______
Teachers		______	______
Teacher Aides		______	______
Other (Specify)	______________	______	______
	______________	______	______
	______________	______	______

11. Number of project participants who received services in each project component area (duplicated count)

Reading		______________
Language Arts		______________
Mathematics		______________
Other Instructional Areas (Specify)	______________	______________
	______________	______________
Supportive Services (Specify)	______________	______________
	______________	______________
	______________	______________

COMPLETE THIS SECTION SEPARATELY FOR EACH PROJECT COMPONENT

12. Number of project participants who received component services by duration of service in months and by last grade completed at the beginning of the project or at commitment if later than the beginning of the project

Grade Duration	3 or under	4	5	6	7	8	9	10	11	Total
0-1										
1-2										
2-3										
3-4										
4-5										
5-6										
6-7										
7-8										
8-9										
9-10										
10-11										
11-12										
Total										

13. State the question(s) addressed for the purpose of evaluating the short-term effectiveness of the component.

Example: Do students who receive reading instructional services from the project make greater gains in reading achievement level than they would have made without that additional instruction?

14. Check the indicator(s) of need for service used to select the participants in the component.

____________ Test Scores
____________ Teacher Judgments
____________ Other Judgments (e.g., self-referral)
____________ Other Performance Indicators (e.g., grades)

15. Check the type(s) of instrument(s) used to measure the intended impact of the component.

____________ Norm-Referenced Achievement Tests
____________ Criterion-Referenced Achievement Tests
____________ Other Tests, Scales and Inventories
____________ Observations
____________ Questionnaires
____________ Interviews
____________ Existing Records

16. On a separate sheet, provide the results of the analysis of the effectiveness of the project component using each of the instruments checked above. Include at least the following:

- a measure of the component participants' average achievement level, performance, etc., after receiving component services (e.g., an average posttest score);
- a measure of the average achievement level, performance, attitude, etc., expected of the participants had they not received component services (e.g., an average pretest score);
- the number of scores the above averages are based upon;
- a brief description/identification of each instrument (e.g., instrument name, edition, level);
- the type of score used in the analysis (e.g., NCE, standard score, raw score, number of objectives, etc.); and
- the date(s) when each instrument was administered or completed.

Appendix 7-B

SAMPLE FORM FOR COLLECTING DATA FOR CHAPTER 1 N OR D PROJECTS

(Short Version)

1. Name of Institution/Facility ____

2. Address of Institution/Facility ____

____

3. Name of Person Completing Form ____

4. Above Person's Telephone Number ____

5. Duration of Project: From ______________________ To ______________________

6. Number of persons under 21 years who resided at the institution/facility during the project (unduplicated count) ______________

7. Number of persons under 21 years who participated in the project (unduplicated count) ______________

8. Number of project staff (paid at least in part by Chapter 1 funds) by job classification (in FTEs)

Admininstrative Staff		________.___
Teachers		________.___
Teacher Aides		________.___
Other (Specify)	____________________	________.___
	____________________	________.___

9. Number of project and non-project staff by job classification who received training funded by the Chapter 1 project

		Project	Non-Project
Administrative Staff		______	______
Teachers		______	______
Teacher Aides		______	______
Other (Specify)	____________________	______	______
	____________________	______	______
	____________________	______	______

10. Number of project participants who received services in each project component area (duplicated count)

Reading		______________
Language Arts		______________
Mathematics		______________
Other Instructional Areas (Specify)	____________________	______________
	____________________	______________
Supportive Services (Specify)		
	____________________	______________
	____________________	______________

COMPLETE THIS SECTION SEPARATELY FOR EACH PROJECT COMPONENT

11. State the question(s) addressed for the purpose of evaluating the short-term effectiveness of the component.

12. Check the indicator(s) of need for service used to select the participants in the component.

__________ Test Scores
__________ Teacher Judgments
__________ Other Judgments (e.g., self-referral)
__________ Other Performance Indicators (e.g., grades)

13. Check the type(s) of instrument(s) used to measure the intended impact of the component.

__________ Norm-Referenced Achievement Tests
__________ Criterion-Referenced Achievement Tests
__________ Other Tests, Scales and Inventories
__________ Observations
__________ Questionnaires
__________ Interviews
__________ Existing Records

14. On a separate sheet, provide the results of the analysis of the effectiveness of the project component using each of the instruments checked above. Include at least the following:

- a measure of the component participants' average achievement level, performance, attitude, etc., after receiving component services (e.g., an average posttest score);
- a measure of the average achievement level, performance, attitude, etc., expected of the participants had they not received component services (e.g., an average pretest score);
- the number of scores the above averages are based upon;
- a brief description/identification of each instrument (e.g., instrument name, edition, level);

- the type of score used in the analysis (e.g., NCE, standard score, raw score, number of objectives, etc.); and
- the date(s) when each instrument was administered or completed.

8

RESOURCES FOR

N OR D PROGRAMS

8. RESOURCES FOR N OR D PROGRAMS

INTRODUCTION

When planning or conducting evaluation acltivities it may be necessary to obtain some outside guidance. There are resources which can be extremely useful to project personnel who need some type of assistance. A primary resource for Chapter 1 programs is that of the Technical Assistance Centers (TACs), which provide consulting services at no direct charge in the area of evaluation. The major focus of this section is to describe the types of services available through the regional TACs. An additional resource which is briefly discussed is the National Diffusion Network, which disseminates information about successful educational programs.

TECHNICAL ASSISTANCE CENTER SERVICES

The United States Department of Education has established regional Technical Assistance Centers (TACs) throughout the country to provide consulting services at no direct charge to projects funded by ECIA, Chapter 1. Depending upon the needs of those being served, TAC services may take a variety of forms, including: (1) on-site visits to deal directly with evaluation problems affecting the project; (2) local workshops dealing with specific issues; (3) telephone consultations to answer specific evaluation questions; and (4) print and mediated materials relevant to Chapter 1 evaluation issues.

Personnel from regional TACs are available to provide help in a variety of topics relevant to Chapter 1 evaluation issues. Some areas of particular interest to N or D projects might include:

- Developing procedures to meet evaluation and reporting requirements;
- Selecting students for a program;
- Developing long-range planning procedures;
- Selecting tests and other instruments;
- Interpreting and using test results;

- Using appropriate data analysis procedures; and
- Applying microcomputers for evaluation and management.

Frequently requested topics for workshops include:

- Selecting Criterion- and Norm-Referenced Tests
- Systematic Teacher Ratings for Needs Assessment
- Developing Composite Scores for Student Selection
- Reporting Evaluation Results
- Conducting Descriptive Evaluations
- Evaluation for Program Improvement
- Developing Tests
- Selecting Measures of Affective Behavior
- Functional Level Testing
- Developing Objectives for Program Evaluation
- Strategies for Program Improvement
- Describing Program Characteristics
- Time-on-Task
- Quality Control
- Sustained Effects
- Test Administration and Scoring
- The Joint Dissemination Review Panel Process
- Evaluating N or D Chapter 1 Programs

For more detailed information on workshops and other consultation services the regional TAC should be consulted. The TAC regions and contractors are listed in Appendix 8-A.

There are some services which are not available through the TACs. These include the enforcement of Chapter 1 rules and regulations, the promotion of particular standardized tests, the gathering of data for studies, the writing of state or local evaluation reports, the recommendation of curriculum or program revisions, and the analysis of evaluation data.

TAC Clearinghouse Resources

The Technical Assistance Center Clearinghouse is located in the Region IV TAC, at the Northwest Regional Educational Laboratory, Portland, Oregon. The Clearinghouse is responsible for collecting, cataloguing and distributing research documents and workshop materials relevant to Chapter 1 evaluation. It should be noted that materials in the Clearinghouse range from subjects of general interest in regard to Chapter 1 evaluation to very specific topics. There are also some items which directly address N or D programs and issues. To obtain more information about the Clearinghouse materials, contact the regional TAC office.

Available Materials

Each regional TAC can also provide copies of various documents relevant to Chapter 1 evaluation, including the evaluation chapters from The Policy Manual, The User's Guide, and The Evaluator's References. The last item is a collection of technical papers, which includes the following titles: (1) An Overview of Hazards to Avoid in the Title I Evaluation and Reporting System; (2) Selecting Students for Title I Evaluation Projects; (3) Composite Scores; (4) Selecting a Norm-Referenced Test; (5) Local Norms; (6) Using Non-Normed Tests in the Title I Evaluation and Reporting System; (7) Out-of-Level Testing; (8) Test Floor and Ceiling Effects; (9) Collecting Achievement Test Data; (10) Factors That Influence Test Results; (11) Problems with Grade Equivalent Scores; (12) Interpreting NCEs; and (13) Score Conversions.

Each regional TAC also maintains audiovisual materials and print packages which, upon request, are available for loan to state, local and project personnel, as well as to other groups. For more specific information regarding print or audiovisual materials, contact the regional TAC office.

THE NATIONAL DIFFUSION NETWORK

The National Diffusion Network (NDN), sponsored by The Education Department, is a program to identify and disseminate successful educational practices. NDN publishes an annual document, "Educational Programs That Work," which describes promising programs. These descriptions, by project, include: (1) characteristics of the target audience; (2) a brief explanation of the program processes and materials; (3) information on evidence of effectiveness; (4) requirements for implementing the program; (5) financial requirements for installing the program; (6) services available from the project to those interested in adopting the program; and (7) a contact person for further information.

Obtaining descriptions about exemplary programs can be useful for both evaluation activities and program planning. Exemplary programs do need to document their effectiveness through some type of sound evaluation procedures. Obviously then, these programs could be a source of information on how to plan and conduct evaluations. Likewise, if student selection procedures are an issue, a review of how a similar program selects students for its program could be a source for new approaches. In the area of program planning, if the evaluation results indicate that improvements need to be made in the curriculum, then programs described by NDN may be a source for new instructional approaches.

Each state has a person who serves as the NDN State Facilitator. These individuals are a good source for the identification of programs which may be of interest and for more information on conferences, newsletters and other publications which focus on exemplary programs. The State Facilitator is also the person to contact for procedures to follow or help in adopting an NDN program.

Appendix 8-A

TECHNICAL ASSISTANCE CENTER REGIONS AND CONTRACTORS

Region I includes: Maine, Vermont, New Hampshire, Massachusetts, Connecticut, Rhode Island, New York, New Jersey, Puerto Rico, Virginia, Delaware, Maryland, the District of Columbia, Pennsylvania, West Virginia, and the Virgin Islands.

The prime contractor is:

RMC Research
400 Lafayette Road
Hampton, NH 03842
(603) 926-8888

The subcontractor is:

Educational Testing Service
Princeton, NJ 08540
(609) 734-5117

Region II includes: Iowa, Illinois, Indiana, Missouri, Wisconsin, Kansas, Ohio, Nebraska, Kentucky, North Dakota, South Dakota, Michigan, and Minnesota.

The prime contractor is:

Advanced Technology, Inc.
1 Park Fletcher Building
2601 Fortune Circle East
Indianapolis, IN 46241
(317) 244-8160

The subcontractor is:

RMC Research
9300 West 110th Street
Overland Park, KA 66210
(913) 341-0008

Region III includes: North Carolina, South Carolina, Georgia, Florida, Mississippi, Alabama, Tennessee, Louisiana, Arkansas, Oklahoma, and Texas.

The prime contractor is:

Educational Testing Service
250 Piedmont Avenue, NE
Suite 2020
Atlanta, GA 30308
(404) 524-4501

The subcontractor is:

Powell Associates, Inc.
3724 Jefferson Street
Suite 205
Austin, TX 78731
(512) 453-7288

Region IV includes: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, New Mexico, Nevada, Oregon, Utah, Washington, and Wyoming.

The prime contractor is:

Northwest Regional Educational
Laboratory
300 SW 6th Avenue
Portland, OR 97204
(503) 295-0214